

AD-A230 076

Technical Report 914

Research and Methods for Simulation Design: State of the Art

Paul J. Sticha

Human Resources Research Organization

①
DTIC FILE COPY

Michael J. Singer

U.S. Army Research Institute

H. Ric Blacksten and John E. Morrison

Human Resources Research Organization

Kenneth D. Cross

Anacapa Sciences, Inc.

September 1990

DTIC
ELECTED
DEC 18 1990
S B D
6



**United States Army Research Institute
for the Behavioral and Social Sciences**

Approved for public release; distribution is unlimited.

90 12 17 101

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS --	
2a. SECURITY CLASSIFICATION AUTHORITY --		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --		4. PERFORMING ORGANIZATION REPORT NUMBER(S) FR-PRD-88-27	
5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 914		6a. NAME OF PERFORMING ORGANIZATION Human Resources Research Organization	
6b. OFFICE SYMBOL (If applicable) --		7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences	
6c. ADDRESS (City, State, and ZIP Code) 1100 South Washington Street Alexandria, VA 22314-4490		7b. ADDRESS (City, State, and ZIP Code) U.S. Naval Training System Center 12350 Research Parkway Orlando, FL 32826-2376	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences		8b. OFFICE SYMBOL (If applicable) PERI-I	
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-85-C-0169	
10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO. 63744A	PROJECT NO. 795
		TASK NO. 345	WORK UNIT ACCESSION NO. C1
11. TITLE (Include Security Classification) Research and Methods for Simulation Design: State of the Art			
12. PERSONAL AUTHOR(S) Sticha, Paul J. (HumRRO); Singer, Michael J. (ARI); Blacksten, H. Ric, and Morrison, John E. (HumRRO); Cross, Kenneth D. (Anacapa)			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 85/01 TO 88/12	14. DATE OF REPORT (Year, Month, Day) 1990, September	15. PAGE COUNT 182
16. SUPPLEMENTARY NOTATION --			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Training system optimization Fidelity Training system model Instructional features Medium selection Training resource allocation	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This report reviews the empirical results and analytical methods available to the training-device designer for tradeoff analyses necessary to product cost-efficient training- device designs. It addresses the problem of training system optimization in three ways. First, it describes existing methods that can aid training-device design functions. The function and operations of methods are compared to the model for the optimization of simulation-based training systems (OSBATS) developed for this project. Second, it reviews research on several issues related to training-device optimization. The issues covered in the review include training-device fidelity, instructional features, skill acquisition, skill retention, transfer of training, and cost estimation. Third, the review organizes the requirements for future research on these topics and sets priorities for research topics based on their cost and the benefit they could offer to the training-device designer.			
(Continued)			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael J. Singer		22b. TELEPHONE (Include Area Code) (407) 380-4387	22c. OFFICE SYMBOL PERI-IF

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

**EDGAR M. JOHNSON
Technical Director**

**JON W. BLADES
COL, IN
Commanding**

Research accomplished under contract
to the Department of the Army

Human Resources Research Organization

Technical review by

John A. Boldovici
Angelo Mirabella

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: HERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22335-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

ARI Technical Report 914

19. ABSTRACT (Continued)

The review focused on quantitative models that can be used to estimate training cost and effectiveness and to determine optimal levels of training-device design variables. The research plan identifies the topics that reduce critical gaps in our knowledge at a reasonable cost. Research addressing (a) relative impact of fidelity features and instructional features on training effectiveness, (b) effects of student aptitudes on training-device design, and (c) organization of nonmonetary reasons for simulation-based training can produce a moderate benefit at a relatively low cost. The most critical research issues involve the impacts of training-device fidelity and instructional features on training effectiveness.

This review provides information that may be used by researchers who wish to develop or improve methods to aid the training-device designers. Designers may use this review to identify methods to aid the training-device design process and individuals who manage research programs may use this information to set priorities for future research efforts.

Accession For	
NTIS GRA&I <input checked="" type="checkbox"/>	
DTIC TAB <input type="checkbox"/>	
Unannounced <input type="checkbox"/>	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Research and Methods for Simulation Design: State of the Art

Paul J. Sticha

Human Resources Research Organization

Michael J. Singer

U.S. Army Research Institute

H. Ric Blacksten and John E. Morrison

Human Resources Research Organization

Kenneth D. Cross

Anacapa Sciences, Inc.

PM TRADE Field Unit at Orlando, Florida

Stephen L. Goldberg, Chief

Training Research Laboratory

Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences

5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel

Department of the Army

September 1990

The cost of training devices and simulators has exceeded, in some cases, the cost of the operational equipment that they service. The capabilities for simulating reality are increasing annually. The problem confronted by the military is to determine exactly how much simulation is sufficient to meet stated learning objectives. Behavioral and analytical techniques that can quickly and easily project or predict how much simulation and training is required are lacking. At the same time, information on variables contributing to cost-effective use of training equipment in courses of instruction is sparse. The development of models, databases, and techniques addressing these problems is the first step toward providing integrated behavioral and engineering decisions in designing, fielding, and using advanced training technology. The potential effect of these tools on the Army is to reduce the cost of training equipment while increasing the equipment's instructional effectiveness.

In response to these concerns and problems, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and the Project Manager for Training Devices (PM TRADE) have joined efforts (MOU of Technical Coordination, May 1983; MOU Establishing the ARI Orlando Field Unit, March 1985; Expanded MOU, July 1986). PM TRADE has maintained partnership in all aspects of the development of the models, databases, and analytical techniques. The final prototype software was delivered to ARI and PM TRADE in December 1988, and has been disseminated to interested parties at Fort Rucker, the Army Training Support Command, and the Systems Training Directorate at the Training and Doctrine Command. The prototype has also been provided to the Naval Training Systems Center Human Factors Research Group, the Air Force Aeronautical Systems Division, the Air Force Human Research Laboratory at Williams AFB, and National Aeronautics and Space Administration Ames Research Center. The models and techniques developed in this effort provide the basis for decision aids that will support the integration of behavioral and engineering data, knowledge, and expertise in training equipment design.

EDGAR M. JOHNSON
Technical Director

FOREWORD

The cost of training devices and simulators has exceeded, in some cases, the cost of the operational equipment that they service. The capabilities for simulating reality are increasing annually. The problem confronted by the military is to determine exactly how much simulation is sufficient to meet stated learning objectives. Behavioral and analytical techniques that can quickly and easily project or predict how much simulation and training is required are lacking. At the same time, information on variables contributing to cost-effective use of training equipment in courses of instruction is sparse. The development of models, databases, and techniques addressing these problems is the first step toward providing integrated behavioral and engineering decisions in designing, fielding, and using advanced training technology. The potential effect of these tools on the Army is to reduce the cost of training equipment while increasing the equipment's instructional effectiveness.

In response to these concerns and problems, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) and the Project Manager for Training Devices (PM TRADE) have joined efforts (MOU of Technical Coordination, May 1983; MOU Establishing the ARI Orlando Field Unit, March 1985; Expanded MOU, July 1986). PM TRADE has maintained partnership in all aspects of the development of the models, databases, and analytical techniques. The final prototype software was delivered to ARI and PM TRADE in December 1988, and has been disseminated to interested parties at Fort Rucker, the Army Training Support Command, and the Systems Training Directorate at the Training and Doctrine Command. The prototype has also been provided to the Naval Training Systems Center Human Factors Research Group, the Air Force Aeronautical Systems Division, the Air Force Human Research Laboratory at Williams AFB, and National Aeronautics and Space Administration Ames Research Center. The models and techniques developed in this effort provide the basis for decision aids that will support the integration of behavior and engineering data, knowledge, and expertise in training equipment design.



EDGAR M. JOHNSON
Technical Director

EXECUTIVE SUMMARY

Requirement:

The goal of this project is to develop methods to help the training-device designer perform the tradeoff analyses required for training-device design. These methods should allow the designer to determine the alternatives that meet training requirements at a minimum cost or provide the maximum training effectiveness at a given cost. The methods should apply to the concept-formulation phase of the training-device development process and should be usable by the engineer responsible for developing the training-device concept. The requirement for this report is to review the empirical results and analytical methods currently available that can be used to support the training-device designer.

Procedure:

This review addresses the problem of training system optimization in three ways. First, it describes existing methods that can aid training-device design functions. The function and operation of these methods are compared to the model for the optimization of simulation-based training systems (OSBATS) developed for this project. Second, it reviews research on several issues related to training-device optimization. The issues that are covered in the review include training-device fidelity, instructional features, skill acquisition, skill retention, transfer of training, and cost estimation. Third, the review organizes the requirements for future research on these topics and sets priorities for research topics based on their cost and the benefit they could offer to the training-device designer.

Findings:

The review focused on quantitative models that can be used to estimate training cost and effectiveness and to determine optimal levels of training-device design variables. The research plan identifies the topics that reduce critical gaps in our

knowledge at a reasonable cost. Research addressing the following three issues can produce a moderate benefit at a relatively low cost: (a) relative impact of fidelity features and instructional features on training effectiveness, (b) effects of student aptitudes on training-device design, and (c) organization of non-monetary reasons for simulation-based training. The most critical research issues involve the impacts of training-device fidelity and instructional features on training effectiveness.

Use of Findings:

This review provides information that may be used by researchers who wish to develop or improve methods to aid the training-device designers. Designers may use this review to identify methods to aid the training-device design process. Finally, individuals who manage research programs may use this information to set priorities for future research efforts.

CONTENTS

	Page
INTRODUCTION	1
Definition of the Scope of OSBATS	1
Organization of This Report	4
TRAINING SYSTEM VARIABLES.	7
Training System Control Variables	7
Training System Operational Variables	8
Other Training System Variables	11
Intervening Variables Defining System Effectiveness.	12
TRAINING SYSTEM OPTIMIZATION MODELS.	15
Overview of OSBATS.	15
Training Effectiveness/Cost Effectiveness Prediction (TECEP)	17
Instructional Systems Development (ISD)	18
The Training Analysis Support Computer System (TASCS)	20
Cost and Training Effectiveness Analysis (CTEA)	22
The Training Developer's Decision Aid and The Training Developer's Decision Support System (TDDA/TDDSS)	24
Computer-Based Task-Sorting Program (TSORT)	26
Cost Effectiveness Methodology for Aircrew Training Devices (CEMATD)	27
Device Effectiveness Forecasting Technique (DEFT)	31
Training Effectiveness and Cost Iterative Technique (TECIT)	32
Summary of Model Functions.	34
TRAINING DEVICE FIDELITY	37
Training Device Fidelity.	37
Transfer-of-Training Literature	55

CONTENTS (Continued)

	Page
INSTRUCTIONAL FEATURES	63
Instructional Features in the Literature.	63
Empirical Research on Instructional Features.	70
Specification of Instructional Features	73
MODELS OF HUMAN LEARNING, RETENTION, AND TRANSFER.	81
Models of Skill Acquisition	81
Models of Retention	93
Models of Transfer of Training.	99
COST ESTIMATION.	109
Cost Modeling: Current Training Systems.	109
Cost Forecasting: Future Systems	111
Estimation Procedures for OSBATS Cost Data.	112
RESEARCH PLAN.	113
Defining the Research Topics.	114
Evaluation of Research Issues	124
REFERENCES	133
APPENDIX A. SYNOPSIS OF FLIGHT SIMULATOR TRANSFER-OF-TRAINING STUDIES.	A-1

LIST OF TABLES

Table 1. Comparison of functions of optimization models	35
2. Characterization of training device fidelity	39
3. Distribution of transfer-of-training studies by aircraft type and stage of training	58
4. Instructional features cited in four sources from the research literature	64
5. Relevant literature for long-range training system cost forecasting methodology.	110

CONTENTS (Continued)

	Page
Table 6. Cost, benefit, and importance weights of research proposals.	126
7. Optimal order of inclusion of research proposals into the research plan	129
8. Optimal selection of research proposals at a cost of \$3 million.	132

LIST OF FIGURES

Figure 1. Interaction of training system components . . .	3
2. Graph of the probability of a correct response by trial for the linear operator model	87
3. Network and matrix representations for the one-element model	89
4. Network and matrix representations of an all-or-none learning model with short- term forgetting	97
5. Research topics and their levels.	117
6. Plot of the benefit and cost of optimal packages of research efforts.	132

RESEARCH AND METHODS FOR SIMULATION DESIGN:

STATE OF THE ART

Introduction

This report reviews the analytical procedures, psychological theory, and empirical findings that form the underpinnings for a model for the Optimization of Simulation-Based Training Systems (OSBATS). The primary goal of the OSBATS program is to provide methods that aid engineers in specifying the set of training devices and concepts for their use that meet training requirements at minimum cost, or provide the greatest training effectiveness at a given cost.

The review presents the basic findings regarding existing training-system optimization models; training device fidelity; instructional features; and psychological models of human learning, retention and transfer processes. We also describe the implications of these results on the OSBATS model. At the conclusion of the report, we summarize the needs for future research and present a plan that identifies high-priority research areas.

Definition of the Scope of OSBATS

The major concern of OSBATS is with the design of simulation-based training systems. In order to communicate clearly the remainder of this review, we first provide definitions that provide the basis for discussing the scope and methods of the OSBATS model.

Definition of a Training System

There are a variety of definitions of the term "training system" that we might use. The definitions vary from broad to quite specific. So that we may reach a satisfactory solution to the problem of training-system optimization, we will be somewhat limited in our definition of a training system. We realize that when we make this definition, the training system that is the concern of the OSBATS model is really a subsystem of a larger system.

We define a training system as a set of activities designed to give students the skills needed to perform operations or maintenance tasks. From this definition, we distinguish the following system components.

1. A target weapon system or job. We are primarily concerned with training in the operation and maintenance of weapon systems, because this is where the potential for the use of training devices is the greatest.
2. A set of training requirements. The training requirements are the activities or tasks that must be performed to set standards at the conclusion of training.

3. Student population. The students being trained are characterized according to their knowledge and skills. We anticipate that different kinds of training would be appropriate for initial skill training, transition training, continuation training, functional training, unit training, and so forth.
4. A trainer. The trainer encompasses both the instructors who deliver the training and the organizational entity responsible for training development.
5. Training methods, Devices, and Simulators. Training methods define training strategies and the use of different training media. Training devices and simulators may be characterized by the extent to which they represent elements of the actual equipment or job environment and the instructional support features they possess.

Figure 1 illustrates how these components interact in defining the training system. The first two components define the controls on the training system considered in the definition. The training requirements specify the criteria for success of the training program. By restricting our attention to training on a single target weapon system, we may deal with single training courses. We are not concerned with problems of allocating training to settings, or with a soldier's career progression through several MOS, although both of those problems have a critical impact on the overall cost and effectiveness of training.

The third component defines the inputs to the training system. The student population characteristics define the scope of the training problem, by specifying the skills of the students who enter training. The scope of the training problem reflects the difference between entering student skills and the skills required after training as specified by the training requirements.

The final two components represent the mechanisms by which the training system operates. Of these components, only the training methods and devices include variables over which we have control in the design of a training system. The OSBATS model is concerned with those variables that are related to training-device design and use. In general, these variables include the fidelity of training devices, the instructional features incorporated in them, and the assignment of training time to training devices. The next section of the report will describe the variables in greater detail.

We judge the optimality of a training system in terms of the cost required to meet the training requirements. In general, we

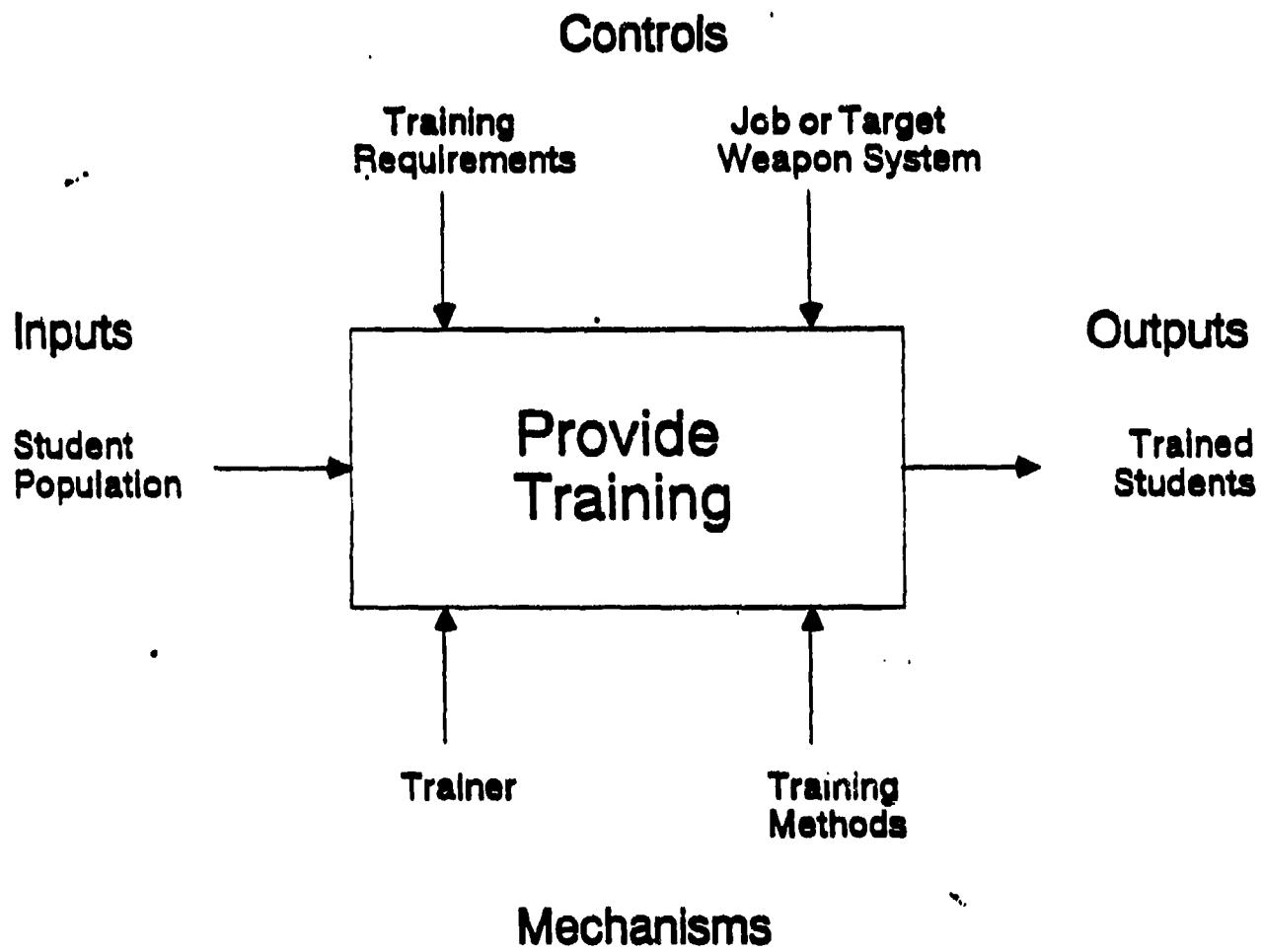


Figure 1. Interaction of training system components.

want to minimize the training cost, while meeting the training requirements.

Support to the Decision Process

Many of the characteristics of a decision aid for training system optimization depend on the stage in the design process in which the decision aid is applied. At earlier stages in the design process, we would expect that less data would be available, and that solutions would be specified in less detail. At later stages in the decision process, the initial solutions would be refined, and additional detail would be added. Given this view of the process, it is critical that the early stages of the decision process weed out bad training-device designs, but less critical that the process determine the best from a set of good designs.

The OSBATS system has been designed to be used in the concept formulation stage of training device development process. At this stage, we are primarily concerned with the functional requirements for a training device, not the engineering specifications. The OSBATS model provides methods for conducting tradeoff analyses that support the determination of the best technical approach to a training-device design.

Organization of This Report

The remainder of this report consists of seven sections. The first section of the report describes some of the critical training system variables in greater detail than was presented above. This description will give additional definition to the scope of the OSBATS model.

The second section describes several general approaches to providing guidance for training system design. We have incorporated parts of some of these approaches into the OSBATS model. Others provide functions that complement those performed by the OSBATS model. Still others have components that may provide alternatives to OSBATS functions.

The third and fourth sections describe two critical components of training-device designs, the specification of appropriate training-device fidelity and the specification of appropriate instructional features. The sections review relevant research in these areas, and describe some of the methods used to select appropriate fidelity levels and instructional features.

The fifth section describes relevant models of human skill learning, retention, and transfer. Ultimately, our ability to optimize the design of training systems depends on our

understanding of how these psychological processes are affected by design variables. Elements of learning and transfer models provide the basis for estimates made by the OSBATS model.

The sixth section describes concerns in cost estimation. It describes methods that characterize the cost of current training devices and methods that could be used to forecast the cost of future training devices.

The final section of the report summarizes the state of the art, and the requirements for further research in the form of a research plan. The research plan lists the areas where research is required to apply the OSBATS model with confidence to a wide variety of tasks. It then organizes the research areas according to estimated costs and benefits. The results of this analysis are used to identify the research areas that have the highest priority.

Training System Variables

Each of the five training-system components identified in the introduction encompasses several variables that both describe the training problem and determine the effectiveness of alternative solutions. Some of these variables are under the control of the training system designer. The OSBATS model is designed to help the designer determine the optimal values for the variables concerned with training methods and devices. Other variables are relevant to the model if one of the following conditions is true: (a) they interact with training methods and devices to affect training cost or effectiveness, or (b) they provide measures of the cost or effectiveness of training methods and device designs.

By restricting our attention to training methods and device variables, we explicitly ignore some training system variables that are relevant to cost and effectiveness. For example, selection of students and instructors, assignment of tasks to training settings, and design of weapon systems are all examples of processes that affect the cost at which training requirements could be met. These processes may have a great impact on the effectiveness of a particular training method or training device design. Indeed, the OSBATS model may be able to shed some light on the effect of changes in these processes. However, the primary analyses of the OSBATS model do not address these processes.

The remainder of this section identifies the principal training system variables. The OSBATS model considers the effects of most of these variables on training cost and effectiveness. However, we have excluded others from consideration, either because they fall outside of the scope of the model or because they would produce more complexity in the model than is warranted in its early stages of development. We first discuss the control variables that define the decisions to be aided by the OSBATS system. Then we discuss other relevant training system variables. Finally, we describe some of the intervening variables that define training system effectiveness.

Training System Control Variables

Training system control variables serve to limit the considerations to be addressed by the modeling effort. The controls used are the job or target weapon system selected and the training requirements for that job or target system.

Weapon System

The weapon system has been used to restrict the scope of the definition of a training system used in this report. The characteristics of the weapon system and of the job that is being trained have a great impact on the training-device requirements.

Weapon system characteristics. The subsystems of the weapon system determine the elements that may be represented in a training device with greater or lesser fidelity. In addition, the existence of certain weapon subsystems may have an impact on the fidelity requirements for other subsystems. For example, if identification of targets is conducted visually, then a device to train target identification must have sufficient resolution to allow for target identification at required ranges. However, if identification is performed using a telescopic sight, then the device must only have sufficient resolution for target detection. The requirement for target detection will be less stringent than the requirement for target identification, depending on the distances involved. The OSBATS currently uses similar reasoning in the rule base that derives task fidelity requirements.

Type of job being trained. Training requirements should be different for different jobs. For example, requirements for maintenance jobs would be expected to produce different types of training devices than operational jobs. Because of its initial focus on a single job, the OSBATS model currently does not consider job characteristics in its analysis.

Training Requirements

Training requirements encompass the performance standard, and other task characteristics. Characteristics of the tasks to be trained are central to the OSBATS model.

Performance standard. The performance standard for a task affects the fidelity requirements. If the performance standard is high, then it will require a training device with higher fidelity to meet the requirements. Similarly, if the fidelity of the training device is held constant, then the required amount of training on actual equipment will increase as the performance standard increases. These considerations form a central aspect of the OSBATS model.

Task characteristics. Task characteristics affect both the required fidelity and the appropriate instructional features. The relationships in this area may be quite specific. That is, tasks that require visual activities require a visual display system; tasks for which motion is a critical cue may require a platform motion system if the motion cue is not correlated with any visual cue, or if the motion cue signals the start of an emergency procedure. This reasoning is currently carried out in the OSBATS model by the rule bases that determine task fidelity requirements and instructional feature appropriateness.

Training System Operational Variables

Training system operational variables include training-device variables and training methods. These variables

encompass the mechanisms that are used to provide training to the student.

Training Device Variables

Producing cost-effective training-device designs is the primary concern of the OSBATS model. The two major classes of training-device variables considered by the model are fidelity and instructional features.

Fidelity. The fidelity of a training device is a measure of similarity between the appearance and operation of the training-device components and the comparable components of the actual equipment. Thus fidelity in itself is not a control variable, but it is a measure of the effect of other control variables. There are many training-device components that can be developed at either a more or less sophisticated level to bring about higher or lower fidelity. For example, the visual system, motion system, and the dynamic simulation system are three such components for a flight trainer.

It is generally assumed that skills learned in a training device with high fidelity will transfer more readily to the actual equipment than skills learned on a device with lower fidelity. (Research regarding this hypothesis is summarized in a later section.) However, devices with higher fidelity are more expensive than devices with lower fidelity, due to their greater technical sophistication. The question that must be addressed in determining the optimal level of fidelity is which specific elements of the training device should replicate the actual equipment with high fidelity, and which may be replicated with low fidelity.

Instructional features. Instructional features are those elements of a training device that allow the instructor to operate the training device, support the instruction, or manage the instructional program. Instructional features can have at least two kinds of benefits: (a) they can have a direct effect on the instructional process to increase training efficiency, or (b) they can have an indirect effect on training efficiency by reducing instructor workload. The OSBATS model is concerned only with those instructional features that have a direct effect on the instructional process.

The concerns in incorporating instructional features into a training-device design are the same as those for fidelity features. That is, the training-device designer must determine which instructional features should be included in the training device, given their cost and the extent to which they improve training efficiency. The training-device designer must also determine how the overall development budget should be allocated between fidelity features and instructional features.

Training Method Variables

The OSBATS model is concerned with training methods chiefly as they relate to the use of training devices. Thus, the major concern of the OSBATS model is the assignment of training to individual devices. There are a host of other method variables outside of the scope of the model.

Training-device use. The variables of concern in this area describe which training devices are used to train each task, and the amount of training that is provided on each device used. Two modules of the OSBATS model address the problem of assigning training time on each task to candidate training devices and actual equipment in order to meet the training requirements at the minimum cost. Ultimately, the assignment of training to devices must consider other constraints on the training system in addition to cost. For example, training on actual equipment may be precluded on some tasks because of safety concerns, or because of the unavailability of appropriate training ranges. In addition, the number of available training devices or pieces of actual equipment may be limited.

Trainer

The trainer includes both the individual instructors delivering the training and the organization responsible for the development and conduct of training. Instructors vary in many ways, including aptitudes, knowledge of the job, experience, and extent of instructor training. The instructor may take on several roles in the training system, including managing the instructional program, providing examples of expert job performance, and giving after-action reviews. The role of the instructor has an impact on the kinds of instructional features that a training device should have. The OSBATS model currently is only concerned with those instructional features that have a direct impact on skill acquisition. While the characteristics of the instructor undoubtedly have a considerable impact on the quality of training, they are outside of the scope of the OSBATS model.

Other Training Method Variables

Media Selection. Methods used to assign training to media other than training devices or actual equipment can have a large effect on the training system. The OSBATS model is exclusively concerned with simulation-based training. However, there are many training media other than training devices and actual equipment that may be used. We assume that appropriate media selection has been completed before the application of the OSBATS model, so that the OSBATS model addresses only the simulation-based component of training.

Training Sequences. Another variable that can have a substantial impact on training effectiveness is the sequence in which individual skills are learned. Although sequence is an important variable, it does not have a great impact on training-device design; consequently, it is not addressed by the OSBATS model. Estimates used by the OSBATS model of the time required to train a task assume that the tasks are taught in a reasonable sequence.

Other Training System Variables

In the following subsections, we list other relevant training-system variables, describe their possible impact on training-system design, and state their current use in the OSBATS model.

Student Population

Students vary in the knowledge that they bring into the training situation, and in the aptitude that they have for obtaining new knowledge. Although student experience is currently considered by the OSBATS model, student aptitude is not.

Student experience. Student experience determines the skills and knowledge that the student brings into the training situation. For example, in transition training, the student generally has some training and experience on similar weapon systems. That is, the student is already proficient on some of the tasks on other equipment. This fact about transition training makes it considerably different in character from initial skill training, where the student possesses few of the required skills. Because of its importance, the entry performance level is one of the primary inputs to the OSBATS model.

Student aptitudes. Student aptitudes may have a variety of effects on training system designs. Some of these effects may be quite subtle. The simplest effect that aptitude may have is that students with higher aptitude learn faster. The implications of this relationship are that the required training would be shorter with higher aptitude students, but the relationship by itself does not have implications on training-device design. It is the more indirect effects of aptitude that have the major implications on training-system design. For example, part-task training strategies may be more appropriate for lower aptitude students. This method-by-aptitude interaction could have a great impact on the design and use of training devices. Because of the complexity of the effects of aptitude, this factor was not considered in the current OSBATS model.

Intervening Variables Defining System Effectiveness

The variables that are used to assess the current state of system performance represent the cost associated with training and the learning processes that define training effectiveness. A brief analysis of what is involved in these two critical performance measures will indicate the kinds of processes that must be understood to make accurate predictions.

Training Effectiveness

The goal of a training system is to provide the students with the skills to perform the complex tasks involved in operating or maintaining a weapon system. Training effectiveness, then, is measured by soldier performance on operational equipment following the completion of the training program. In order to predict training effectiveness, one needs to know how criterion performance in the training environment transfers to performance in the operational environment. Knowing the extent of transfer of training, one could determine the training criterion on a simulator that would produce criterion performance on operational equipment. Similarly, it would be possible to predict the operational performance resulting from any amount of training on a training device.

Training effectiveness, then, is affected by two variables: (a) performance criteria on training devices, and (b) transfer of training from training devices to operational equipment. The performance criteria are control variables specified in the training system design. Transfer of training is the major state variable involved in assessing training effectiveness. To develop procedures that optimize training-system design, we must understand transfer of training as it relates to the training-system control variables that describe training-device design options.

Training-System Cost

The life-cycle cost of a training system may be divided into two major components: (a) one-time development and procurement costs, and (b) ongoing operating costs. The one-time development and procurement costs are a function primarily of the complexity of the training equipment and are relatively independent of how the equipment is used. A model of training-system cost, therefore, must be able to predict procurement costs as a function of equipment sophistication and complexity.

Ongoing operating cost is a function of both the complexity of the training system and the extent to which the system is used. The extent to which the system is used, in turn, is affected by the characteristics of skill acquisition and retention processes, as they relate to the instructional features

and technical characteristics of the training devices and the requirements of the tasks to be trained. To the extent that training time may be reduced by proper training-device design or sequencing of training, the operating cost of the training system may be reduced. For training in an institutional setting, the characteristics of skill acquisition and the effects of sequencing of training are of primary importance. For unit training, skill retention also plays an important role in determining the requirements for system use, and hence its cost.

Training System Optimization Models

This section first describes the structure and concepts of the OSBATS prototype and then reviews existing models that also address the problem of optimizing training-system design. Some of these methods, such as the Interservice Procedures for Instructional Systems Development (ISD; Branson, Rayner, Cox, Furman, King, and Hannum, 1975) are much more general than the OSBATS model. Others, such as the Cost Effectiveness Methodology for Aircrew Training Devices (CEMATD) (Marcus, Patterson, Bennett, and Gershan, 1980), serve a very similar function to components of the OSBATS model. Nine of these methods are reviewed. Where the OSBATS model has used concepts from other methods, we have described the relationships between the methods.

Overview of OSBATS

A detailed description of the OSBATS model is presented by Sticha, Blacksten, Buedo, Singer, Gilligan, Mumaw, and Morrison (1990). Summaries of the model are available from several sources (Sticha, Blacksten, and Buedo, 1986; Singer and Sticha, 1987; Sticha, 1989). The prototype OSBATS model currently consists of the following five modeling components.

1. **Simulation Configuration Module.** A tool that clusters tasks into the categories of part-mission training devices, full-mission simulators, and actual equipment.
2. **Instructional Feature Selection Module.** A tool that analyzes the instructional features needed for a task cluster and specifies the optimal order for selection of instructional features.
3. **Fidelity Optimization Module.** A tool that analyzes the set of fidelity dimensions and levels for a task cluster and specifies the optimal order for incorporation of advanced levels of these dimensions.
4. **Training Device Selection Module.** A tool that aids in determining the most efficient family of training devices for the entire task group, given the training device fidelity and instructional feature specifications developed in the previous modules.
5. **Resource Allocation Module.** A tool that aids in determining the optimal allocation of training time and number of training devices needed in the recommended family of training devices.

The concept of operation for the OSBATS model is based on the iterative use of the five model tools to make recommendations regarding the definition of task clusters, the design of training

devices, and the allocation of training resources among selected training devices. Both the subset of tools that are used and the order in which they are used may vary depending on the requirements of the problem and the preferences of the user. Although the tools may be used in a variety of orders, the most natural order is the order in which the tools were listed above. An application of the tools in that order is described in the following text.

In this example, the analyst uses the Simulation Configuration Module to examine the tasks to be trained and to provide a preliminary recommendation for the use of either actual equipment or one or more training devices. The result of this analysis is three clusters of tasks. Two of these clusters define tasks for which a full-mission simulator or part-mission training device should be designed.

The analyst then uses the task clusters defined by the Simulation Configuration Module as the basis for the application of the Instructional Feature Selection and Fidelity Optimization Modules. These two modules define candidate training-device designs for each task cluster. The output of the two modules is a range of options that vary in cost. Thus, the overall results of the application of these modules is a collection of training device designs specifying for each design the level of fidelity on each fidelity dimension and the collection of instructional features included in the design. The analyst selects several of these designs for further examination.

The Training Device Selection Module evaluates the training device design produced in the previous process. The analyst exercises this module several times using different combinations of training devices. For each combination, the module determines the number of tasks assigned to each training device, the number of hours each task is assigned to each device to meet the training requirements at the lowest cost, and the optimal training cost given the particular combination of training devices. This model makes the simplifying assumptions that the hourly cost of a training device is fixed and that all devices are fully utilized. These assumptions allow the Training Device Selection Module to determine a solution in less than one minute.

When the analyst is relatively confident of the solution of the Training Device Selection Module, he or she then investigates the solution using the Resource Allocation Module. It could be that the recommendations of the Training Device Selection Module would require the procurement of more training devices than are feasible, or would recommend training on actual equipment for tasks in which such training violated safety regulations. The Resource Allocation Module allows the analyst to impose constraints such as these on the training system and examine the resulting optimal solution. The Resource Allocation Module also

relaxes the simplifying assumptions that were used by the Training Device Selection Module to estimate training device cost, leading to a more accurate cost function. As a result of its increased generality, the Resource Allocation Module takes several minutes to reach a solution, several times longer than the Training Device Selection Module.

At many points in the analysis process, the analyst has the option of returning to modules that were used previously to refine the analysis, change assumptions, or choose different solutions. For example, the analyst might change the definition of the task clusters based on the results of Training Device Selection Module, or may use those results to select different candidate device designs for evaluation.

Training Effectiveness/Cost Effectiveness Prediction (TECEP)

The prototype method for media selection is contained in the Training Effectiveness Cost Effectiveness Prediction (TECEP) methodology (Braby, Henry, Parrish, and Swope, 1975). TECEP methods have been incorporated into several other models including the Interservice Procedures for Instructional Systems Development (ISD), the Automated Instructional Media Selection (AIMS) procedures (Kribs, Simpson, and Mark, 1983), and versions of Cost Training Effectiveness Analysis (CTEA). In addition, the OSBATS model has used some of the concepts from TECEP to predict the effectiveness of instructional features.

The first step in the TECEP media selection process is to classify training objectives according to the type of information-processing activities required in each task. TECEP considers the following twelve classes of tasks:

1. Recalling bodies of knowledge
2. Using verbal information
3. Rule learning and using
4. Making decisions
5. Detecting
6. Classifying
7. Identifying symbols
8. Voice communicating
9. Recalling procedures, positioning movement
10. Steering and guiding, continuous movement
11. Performing gross motor skills
12. Attitude learning

Each training objective is associated with a learning algorithm. The learning algorithm is "a step-by-step prescription for a student to follow in learning any specific task in a class of learning tasks. . . a general sequence for use with all similar training objectives" (Braby, et al., 1975, p. 14).

Each learning algorithm is associated with a set of appropriate media. Instructional media are selected from this list according to their capability to provide the essential stimulus characteristics to allow the trainee to respond to them, and provide feedback and reinforcement. Each task category has a chart to be used for media selection; the chart lists the potential media and criteria for selection and indicates which media meet which selection criteria. The user determines which criteria must be met and selects the instructional medium or media that meet all relevant selection criteria. More recent adaptations of this method, such as AIMS and CTEA, have replaced the dichotomous criteria with numbers that assess the extent to which the selection criteria are met. In these cases, an overall measure of the capability of each training medium may be made.

The OSBATS model uses procedures based on TECEP to determine the appropriateness of instructional features for individual tasks. The tasks are classified according to the taxonomy described above, learning algorithms are determined using TECEP procedures, and instructional features that are required to support the learning algorithms are recommended. The relationships between task categories and instructional features determined by the analysis are summarized in a rule base that relates instructional features directly to task characteristics. The rule base combines the analysis based on TECEP with other analyses.

Instructional Systems Development (ISD)

The media selection guidelines of the ISD model have been adopted for use by instructional developers of U.S. military training to aid in the selection of instructional media for training systems (TRADOC Pam 530-30, 1975; TRADOC Reg 350-7, 1981). The media selection process is accomplished using flowcharts to aid the user in the task. The flowcharts help determine which forms to obtain and how to gather and correctly place information on the data forms. The first step in the ISD process is to complete a Delivery System Planning sheet where the user must:

1. Determine the selected delivery system (e.g., simulator with adjunct displays),
2. Provide a rationale for his/her choice,
3. State the learning objectives, and
4. Complete a learning category matrix indicating the extent to which the learning objective requires gross motor skills, classifying, or attitude learning.

From this point any unavailable techniques are eliminated and the remaining systems are compared on the basis of the learning category criteria (complexity, administrative, stimulus, etc.). After narrowing down the alternatives to one or more systems, the user then selects the most likely medium presentation system from the list of candidates. A system may be rejected due to one or more of the following reasons: size, interface, time to produce, budget, or an under-developed state of the art.

Two studies have examined the use and usefulness of ISD media selection procedures. In the first such study, Vineberg and Joyner (1980) examined the instructional development process of 57 courses sampled from Army, Navy, Air Force, and Marine Corps offerings. They found that there were only three instances (about 5% of the total number) where instructional developers attempted to select media according to the ISD procedure. In the remaining 95%, developers did not even attempt to select media at all. The reason most often reported by developers was that they were not free to change the media that existed for instruction. In a few cases, a higher command actually dictated the use of a new medium or media that presumably offered certain advantages over the existing media; the developers task was to redesign the course to integrate these new media. In these cases, the media were selected before the ISD procedure started.

Vineberg and Joyner (1980) also pointed out that the ISD media selection procedure is flawed in that it depends on a previous step in the ISD process, specification of instructional activities. The specification process is itself not well developed. Guidance is "...provided largely by example rather than by means of explicit decisional rules" (p. 98). They concluded that training developers should not be expected to select media according to ISD procedures until the prerequisite procedure for specifying instructional activities is improved.

In the second study of the ISD media selection procedure, Gagne, Reiser, and Larsen (1981) informally surveyed 29 instructional developers at four Army schools. The researchers found that more than 50% of the developers considered it preferable to have new media selection guidelines developed, while 8% of the developers believed that the ISD guidelines should be revised. Many of the instructional developers believed that certain portions of the ISD guidelines were not specific enough, while at the same time other portions of the guidelines were too detailed. It was stated that many of the terms used in the guidelines needed to be more adequately defined, and that more examples were necessary. Some of the developers considered that too many learning categories were used, and that, in general, the guidelines could be simplified and condensed. Indeed, Montemerlo (1975) suggests that the ISD model does not provide sufficient guidance for the novice, and is primarily useful to the training developer who is already expert.

The Training Analysis Support Computer System (TASCS)

The Training Analysis Support Computer System (TASCS) (Plaats, Butler, Hays, and Atkins, 1982; Logicon, Inc., 1982) is an automated media selection model designed to aid training developers in the instructional systems development process. In general, the TASCS process begins with completed task statements generated earlier in the development process, transfers these task statements to objective statements, selects appropriate media to accomplish the objectives, and generates a course syllabus according to the objectives and media selected. The entire process is divided into five distinct phases. A description of each phase follows.

Task Analysis

During the task analysis phase, task data generated prior to TASCS are entered into the system. When an entry is complete, the user is prompted to make selections of how each task statement should be characterized, e.g., criticality of tasks, etc. Task record numbers and performance statements are then printed to aid the user in developing a task hierarchy. The task hierarchies are developed manually by assigning hierarchy numbers to each task and entering the numbers into the system.

Objective Analysis

The ordered task statements developed in the first phase are transferred in the objective analysis phase to objective data records. Each objective statement is examined to determine if it is stated correctly. A correct statement is one that includes the conditions that the training will occur under, the standards that must be met, and the actions that must be performed. In addition, each objective is assigned to at least one Learning Subcategory (LSC). LSCs are found in the Instructional Quality Inventory (Naval Personnel Research and Development Center, 1979) and classify the objectives on two dimensions: task level (remember, use unaided, use aided) and content type (fact, category, procedure, rule, principle). The objective statements and their associated LSCs are then ordered in a hierarchical fashion similar to the procedure described for ordering the task statements. This hierarchy serves to "...depict the learning relationships between the objectives and to establish prerequisite skills and knowledges which are needed prior to advancing to the more complex or integrated performances" (Plaats, et al., 1982).

Media Analysis

The concept of the Learning Experience comes into play for the media analysis phase. A learning experience is the vehicle that is used by the trainer to present the lesson or course

material to the student and is used in the TASCS to provide a "common denominator" between objective statements and media. TASCS recognizes eight learning experiences that are derived from the ISD model. These are, explanations (dynamic, graphic, and textual), demonstrations, part-task practice and test (cognitive, psychomotor, and affective), and full practice and test.

Along with these learning experiences, 19 media are identified and included in the TASCS. Each medium is rated in terms of its instructional capability, administrative capability, and cost. When the user is satisfied that the media have been accurately represented, the "media pool" is assigned to the learning experiences. A minimum level of performance is specified by the user, and the system responds with a listing of the media that can be applied to that learning experience given the performance level. If a new performance level is entered, a different set of media for that learning experience is displayed.

Instructional Analysis

The instructional analysis phase has three major subgoals.

1. It assigns the appropriate learning experiences to each objective in accordance with the characteristics of that objective. The salient characteristics in this step are the objectives learning subcategory (LSC), the difficulty and criticality rating, and the reason for the difficulty of the performance.
2. It assigns an evaluation methodology to each objective that is consistent with the learning subcategory assigned to that objective.
3. Specific media are assigned by "...ranking the media associated with the objectives Learning Experience in order of either instructional costs and/or administrative capabilities" (Plaats, et al., 1982). In other words, each medium set assigned to each learning experience for each objective is ranked with regards to special requirements necessary for that learning experience.

TASCS will then provide "possible solutions," not just one answer, to meeting the objective. The trainer or training developer then selects a solution.

Syllabus Development

The final phase of TASCS is to print course syllabus outline for use by the training developer. This syllabus defines the "...sequence of objectives within a course/week/day/hour, and

a listing that details all characterizations for each objective in presentation order for use as a lesson specification" (Plaats, et al., 1982). Course syllabus development requires that the user input information concerning the grouping preferences of the analyst, the identification of time to train each objective, and identification of any constraints that exist.

Cost and Training Effectiveness Analysis (CTEA)

The Cost and Training Effectiveness Analysis (CTEA) is the first of two methods described in this section that apply the general optimization techniques known as Multiattribute Utility Measurement (MAUM) to the media selection process. MAUM methods have been applied to a variety of decision problems involving the selection of a single alternative from a set of candidate alternatives that are characteristically multidimensional. Hogarth (1980) states that MAUM techniques can be characterized by a basic set of features. These features include the selection of dimensions for evaluation, the determination of adequacy on each dimension, the derivation of comparable measurement scales across dimensions, the weighting of and aggregation across dimensions, and the selection of the outcome or alternative with the greatest score. Hawley and Dawdy (1981) describe the objective of the MAUM concept as follows:

Every outcome of an action has a value or utility on a number of different attributes, dimensions, or factors. The objective of MAUM, in any of its numerous versions, is to determine these values, one factor at a time and then to combine them across factors using a suitable aggregation rule. (p. 1-5)

Cost and Training Effectiveness Analysis (CTEA) is a methodology that has as its primary goal, "the optimization of soldier capability at a minimum cost" (Dawdy, Chapman, and Frederickson, 1981a). CTEA is a 12-step process ranging from the development of a detailed research plan and the identification of medium/method sets, to comparing relative training cost and effectiveness measures and recommending training programs with the "best" cost and training effectiveness. The process is organized into two primary analyses: a training effectiveness analysis and a cost analysis. The training effectiveness analysis focuses on identifying "...the percent of the described students that can be expected to reach 100% of the training standard..." and the cost analysis is directed toward identifying "...cost elements inherent in the acquisition and operation of any training program" (Dawdy, et al., 1981a).

In the first step of the CTEA process, task lists are developed and evaluated in terms of their criticality. In applying this process, each task is rated either High, Medium, or Low on the following nine dimensions:

1. Difficulty of learning,
2. Difficulty of performance,
3. Importance to mission success,
4. Importance to personal survival,
5. Frequency of performance,
6. Peacetime performance requirements,
7. Wartime performance requirements,
8. Elapsed time between task cue and performance, and
9. Consequence of inadequate performance.

The ratings are made by subject-matter experts (SMEs); tasks are selected for training according to these ratings. Formal task statements are developed to identify the standards to be achieved, the conditions under which the task must be performed, and the task-enabling skills required.

The tasks/skills are further identified in terms of their interdependencies (e.g., Does successful learning of task B depend on successful learning of task A? Is task B a logical subset of task A?) and then ordered into a training sequence. The ordering is based on one of three rules (i.e., order tasks from simple to complex, order tasks in sequence in which they are performed, or order functional groupings of tasks) with the selection of a particular rule dependent on the nature of the task for which training is being developed.

The process continues to identify possible methods/media for training each task. The approach for selecting these media is the TECEP method described previously. Briefly, the TECEP method involves classifying tasks in terms of learning algorithms, and analyzing these tasks to "determine the characteristics of the stimuli that control task performances" (Dawdy, Chapman, and Frederickson, 1981b).

The methods/media are then consolidated for all tasks to form a quasi Program of Instruction (POI). These POIs serve two functions: They provide a framework for collecting training effectiveness data and training design information essential for developing training courses, and they provide the SMEs with an understandable format to be used as a basis for assessing training-alternative effectiveness. With the information provided up to this point, the training program options are developed by SMEs.

After the options have been developed, a costing procedure is employed that focuses on identifying the particular cost elements associated with each training program option. The costs are partitioned to reflect costing variation sources between training options and to indicate sources of funding and a projected schedule of expenditures.

A MAUM-based forecasting method is used to obtain a measure of training effectiveness.

Using utility-theory-based scaling procedures, the worth of training specific tasks and the effectiveness of training the same tasks were combined to yield quantitative measures of training program worth. An overall measure of training worth for each task was obtained by summing the partial measure of training worth across tasks to obtain the appropriate training worth weights. The aggregate measure of training worth obtained in this fashion was quantitative and thus suitable for providing a foundation for establishing the cost-effectiveness comparison (Dawdy, et al., 1981b, p. 2-48).

Methods based on MAUM have been used in OSBATS, especially in the Simulation Configuration Module. Their use in that module is similar to the process used in CTEA to select tasks for training. However, in OSBATS the tasks are selected for training on a single training device, which may be either a full-mission simulator or a part-mission simulator.

The Training Developer's Decision Aid and the Training Developer's Decision Support System (TDDA/TDDSS)

The Training Developer's Decision Aid (TDDA) (Frederickson, Hawley, and Whitmore, 1983) and the Training Developer's Decision Support System (TDDSS) (Hawley and Frederickson, 1983) automate portions of the CTEA methodology and aid in developing training programs. TDDA provides support to the CTEA process during the Training Design phase, including a function analysis, task analysis, and learning requirements analysis. The end product of TDDA is a set of alternative training media. TDDSS goes a step further in the CTEA process to include the Training Evaluation phase. This includes the resource projection, cost estimation, benefit analysis, cost benefit integration, and alternative selection processes of the CTEA methodology.

Four major changes from the CTEA framework were made in the development of TDDA/TDDSS. The CTEA design process begins with the identification of media/methods to be used in the training process. The emphasis is on training delivery rather than on acquisition of skills or knowledge by the learner. TDDA/TDDSS, however, attempts to change this by placing the emphasis on the acquisition of these essential skills and knowledge through the specification of Functional Learning Requirements (FLR). The eight FLRs are:

1. Set the learning objective.
2. Establish the performance context (i.e., cues and consequences of inadequate performance).
3. Provide performance instructions.
4. Demonstrate the performance.
5. Provide practice situations for each task.
6. Provide performance feedback.
7. Provide corrective guidance.
8. Establish the appropriate level of understanding for the materials presented.

The inputs that are given for each of these requirements form the basis for prescribing a delivery system, including specifications for the media to be used in training, that will enhance instructional quality.

The second major difference between CTEA and TDDA/TDDSS is in the automation of the systems. TDDA/TDDSS are programmed and implemented on Apple II+ computers. This change has resulted in the following three improvements to the training development process. (a) It aids in the acquisition of information needed for the instructional development procedures; (b) it organizes this information into databases; and (c) it provides database management and analysis capabilities to support the procedures.

The third change was in the use of Expert Job Performers (EJPs) as opposed to Subject Matter Experts (SMEs) to develop the tasks lists that drive the TDDA/TDDSS process. An EJP is defined as an individual who has had at least 18 months of direct operational job assignment experience in a primary job position within that assignment, with the experience occurring in the last three years. By using EJPs it is hoped that more accurate task lists will be generated.

The fourth change was the development of three modules for job analysis, capable of generating task lists from task descriptions provided by the Expert Job Performers (EJP). These modules are differentiated by job type and were designed because of the need to describe a job in terms of its structure (i.e., the object upon which work is focused and the work behavior that surrounds it). The job modules are the maintenance job module, the equipment use/operation job module, and the information/data processing job module.

TDDA/TDDSS has been successfully applied to the Remotely Monitored Battle Field Sensor System (REMBASS), the aircrew positions on the AH-64 Apache Attack Helicopter, and training design support for the Patriot Engagement Control Station Operators (Hawley and Frederickson, 1983).

Computer-Based Task-Sorting Program (TSORT)

TSORT is an automated method designed to aid nuclear-power-plant training analysts in determining if tasks are being trained by appropriate training media or strategies. More specifically, it "...provides a standardized method to select tasks for use in Nuclear Regulatory Commission (NRC) training research..." and it assists in evaluating "...whether training program developers have allocated nuclear power plant tasks to appropriate training strategies" (Jorgensen, 1984). The methods used by TSORT are based upon those employed in the early phases of CTEA.

Training analysts provide the primary input to TSORT in the form of values for the following ten dimension for each task:

1. Skill acquisition difficulty,
2. Skill performance difficulty,
3. Immediate performance need,
4. Safety consequences,
5. Previous nuclear experience,
6. Normal operation performance,
7. Emergency operation performance,
8. Plant delay tolerance,
9. Regulatory requirement, and
10. Economic consequence.

These values are input into the computer through menu screens and system prompts.

A particular value, or range of values for a dimension is associated with one or more of a given set of training strategies or categories. That is, a specific criterion level must be met on a dimension for consideration of a particular training strategy. The following training categories are identified:

1. Qualification training,
2. Certification training,
3. Refresher training,
4. Elimination candidate (eliminate training for the task),
5. On-the-job training,
6. Candidate for less training,
7. Candidate for more training,
8. Candidate for simulation training, and
9. Candidate for formal training.

TSORT has the capability of using two different types of metrics. The first type, a count metric (absolute value), is used when the emphasis is on selecting a training strategy for individual tasks. For each of the training strategies, the number of dimension values that meet the criterion for that strategy is counted. This metric represents the total number of task ratings that fall within an acceptable range.

The second type of metric was developed for TSORT to indicate how far from the criterion level a certain dimension value was. Each dimension is coded with a number that indicates the direction and magnitude of deviations from the criterion. The values are then averaged with the resulting value providing a means for rank ordering.

After the value dimensions have been entered, the user may then analyze the data by sorting and ranking them by either their "match values" (count metric), or the "average values" (relative value). An additional option allows the user to look at a ranking of the tasks for any particular dimension. For example, if "...a rank ordered list of skill acquisition difficulty on tasks..." is selected, the computer will generate a rank ordered list of the tasks in terms of their skill acquisition difficulty. Finally, the user may perform a cost-benefit analysis. The user inputs cost information concerning the operating cost of the nuclear plant in terms of the tasks performed and the plant environment in general. The computer then generates a rank ordered list of tasks in terms of their "dollar cost of poor training."

Jorgensen (1984) suggests that further uses of TSORT might include the ranking of training scenarios rather than tasks and that any application of TSORT "should be based upon carefully agreed upon criteria and dimensions."

Cost Effectiveness Methodology for Aircr Training Devices (CEMATD)

The Cost Effectiveness Methodology for Aircr Training Devices (CEMATD) is intended to be an automated cost-benefit model to allocate training on tasks to instructional media in such a way as to satisfy several training objectives at minimum cost (Marcus, et al., 1980). Developed for the Air Force Human Resources Laboratory (AFHRL), the model currently is not being used. It was shelved after failing to exhibit sensitivity to parameters in a logical way. According to AFHRL, the modelers "were never able to get the interaction between cost and training effectiveness."

We found the documentation difficult to understand and believe this is due to formulation problems lying at the root of the model's failure. This criticism and the apparent failure of

the model notwithstanding, the study was an ambitious attempt to solve a very difficult problem, and the report contains information of possible value.

The modeling approach attempts to consider several training objectives simultaneously. Its procedures have the following characteristics.

1. It assumes that the amount of per-student training time on a device--if the device is used at all--is known a priori. It thus assumes that the transfer of training, as measured by the cumulative transfer effectiveness ratio (CTER) does not vary with the amount of training on a training device. This assumption may lie at the root of the problem with the model. Determining the amount of per-student training time on a device is at the center of the cost-effectiveness issue, as the preceding authors have argued. Further, the CTER is a strong function of training time on a device.
2. It specifically addresses costs associated with the number of training devices procured.
3. It determines the optimal number of training devices of various types to procure by enumeration of all possible solutions, rather than by analytical optimization techniques. This was possible due to assumption 1, above.

The CEMATD model was divided into six processes, (a) input processing, (b) generation of alternatives, (c) determination of capabilities, (d) determination of effectiveness, (e) calculation of cost, and (f) output processing. A description of each process follows.

Input processing

The input process provides the basic information to drive the model. This information is categorized into one of three categories.

1. Training requirements data. This category includes information about the number of training components involved (i.e., tasks), the average number of hours required in each device to train to criterion, and the number of aircrew trainees to be trained for each level of training.
2. Training device data. This category includes information about the capability of the device to satisfy the training requirements of each task, the maximum amount of time that the training device would be available for the training program (in hours), and the number of training bases.

3. Training cost data. This category includes procurement costs, operating and support costs, the economic lifetime of the device, and discount and inflation factors.

The model gives the user the following options to specify the type of input appropriate to the training situation.

1. The ability to input data for either functional or mission-related tasks,
2. The ability to enter functional requirements as a function of the devices,
3. The ability to express device capabilities in terms of total requirements or as transferable requirements only,
4. The option to enter costs from available "cost experience," or from "cost-estimating relationships,"
5. The option to include or exclude of TDY,
6. The option to express TDY in terms of "the number of trips each year," or "the number of days of TDY per trip,"
7. The ability to vary data items independently for sensitivity analysis, and
8. The capability to examine escalated costs or non-escalated costs.

Generate alternatives/determine capabilities

The model uses an algorithm that generates all the possible combinations of devices. These alternatives are then analyzed to determine which ones have the highest capability. Capability in this model is defined as the ability of each device to meet the training requirement, and the availability of the device. It is assumed that devices for tasks that are not unique can be "nested." That is, a device with the highest capability can meet all the requirements of the next highest capability, etc., providing data that specify "the maximum design performance of each individual device for each task without respect to other devices being evaluated."

Determine effectiveness

The alternatives derived in the previous stage are then examined to identify their effectiveness. Alternative effectiveness is described as the satisfaction of all the training task requirements. An effectiveness measure is derived in three steps. First, the devices are ordered in terms of their capability (described in equivalent aircraft hours per training-

device hour). Then a time (in hours) is assigned to each device to satisfy total crew requirements for that task. The hours are then summed for each device and compared to the maximum availability of the device. If the alternative fully meets the training requirements and the total time required of the alternative does not exceed its availability, it is further analyzed in terms of cost: otherwise, it is discarded.

Cost processing

There are six major cost components input to the model. These components are: (a) acquisition costs, (b) operation cost, (c) base operating support costs, (d) logistics support cost, (e) personnel support costs, and (f) recurring investment costs. A life-cycle cost analysis is used based on the Air Force procedure for calculating such costs. Basically, the costs associated with procurement of an alternative (acquisition costs X number of devices) are combined with the total operating costs (Hours trained X direct operating and support costs and TDY costs) with consideration made for discount factors and inflation rates, to arrive at a total training cost figure.

Output processing

A matrix is then developed to match the candidate set of training alternatives with their associated life cycle costs. Other outputs include a summary of the costs for the most effective alternatives, a cost breakdown, and a utilization breakdown of each device type in each alternative. These breakdowns allow the user to perform a sensitivity analysis. The final selection of an alternative is done by selecting one of the alternatives from the remaining set of alternatives.

In summary, the CEMATD model is a complex model that somehow failed in its objective. Its very complexity, and the lack of a clear statement of all assumptions, makes it difficult to pinpoint the problems in formulation. Nevertheless, the report contains some interesting concepts and a good listing of cost components for consideration.

As this description indicates, the goals of CEMATD are similar to the goals of the Training Device Selection and Resource Allocation Modules of OSBATS. Both models are concerned with allocating training time to training devices that differ both in their cost and in the extent to which training on the devices transfers to actual equipment. Despite this similarity in goals, there are many differences in the methods used by these OSBATS modules and those used by CEMATD. Most notably, the CEMATD model assumes that both the transfer of training and the required training time can be estimated by the user for any training device, while the OSBATS model estimates these values by

comparing the training-device fidelity and instructional features to the task requirements.

Device Effectiveness Forecasting Technique (DEFT)

The Device Effectiveness Forecasting Technique (DEFT; Rose, Wheaton, and Yates, 1985) is based on a program-evaluation framework that evaluates a training device as an element of the training system of which it is a component. The program evaluation model used by DEFT considers several training activities, including (a) preliminary training such as classroom training, (c) device-based training, and (c) actual equipment training. The model also defines the inputs, and intermediate and terminal outputs of the training system.

The DEFT model consists of the following four activities.

1. Training Problem. The assessment of the magnitude of the training problem considers both the difference between the input skills of the students and the performance standard and the difficulty of training to meet the performance standard.
2. Acquisition Efficiency. This factor measures the effectiveness of the training conducted on the training device. The model assesses acquisition efficiency based on the training principles and instructional features used by the training device.
3. Transfer Problem Analysis. This analysis addresses the magnitude of the training problem that remains following training on the device. The analysis is based, in part, on the fidelity of the training device.
4. Transfer Efficiency Analysis. This analysis is concerned with how well the skills learned on the training device transfer to the actual equipment. The analysis is based on device principles that aid transfer of training.

The model combines the ratings algebraically to estimate training-device effectiveness.

There are three levels of DEFT that operate at different degrees of detail. DEFT I is the least detailed, and can obtain an effectiveness estimate based on global judgments. DEFT II bases its estimate on task-level judgments. DEFT III is the most detailed, operating at the subtask level.

There is some similarity between DEFT and OSBATS in that both models recognize the importance of evaluating a training-device within the context of the training system in which it is embedded. In this respect, OSBATS is much more complete and flexible than DEFT, in that OSBATS addresses situations with

multiple training devices, and estimates the training time and cost required to meet training requirements. The major distinction between DEFT and OSBATS is that OSBATS is developed as a design tool rather than an evaluation tool. Consequently, the OSBATS model gives the user the capability to investigate many design alternatives simultaneously, while DEFT requires the user to evaluate alternatives sequentially.

Training Effectiveness and Cost Iterative Technique (TECIT)

TECIT is a model that evaluates the cost effectiveness of a training device or simulator (Goldberg, 1988). It was designed to be used at several stages in training equipment development cycle. At early stages in the development process, before a training device has been produced, the results of the model are based on estimates made by the analyst or by subject matter experts (SMEs). When the device has been fielded, empirical data may replace or supplement the analytical estimates.

The TECIT model may address the following questions concerned with designing training devices, forecasting training-device performance, and validating the model:

1. Determining whether a training device or simulator should be developed,
2. Selecting the best training device design from competing design alternatives,
3. Optimizing the cost effectiveness of a training device design,
4. Evaluating a device design for acceptance testing,
5. Forecasting skill acquisition using a training device,
6. Forecasting transfer of training using a training device,
7. Forecasting training deployment and time,
8. Designing empirical studies of acquisition learning and transfer of training,
9. Designing empirical studies to validate the model.

TECIT evaluates the effectiveness of a training device or simulator considering safety, skill acquisition, transfer of training, and device utilization using the following relationship.

$$TD/S\ E\ (f) = \frac{S, ToT, JR}{Acq} UR$$

where

TD/S E (f) denotes the training device effectiveness function,

Acq = the acquisition learning on the device measured in terms of time to criterion,

S = a safety rating,

ToT = transfer of training from the device to an exercise on the weapon system during training,

JR = a rating of job readiness for a work sample device (or the transfer of training from the device to the job), and

UR = the utilization ratio of the device (the proportion of scheduled hours are actually used)

The three factors in the numerator of the function are combined using a weighted sum. The weights are based on the judgments of the analyst of the importance of the three factors.

The TECIT report describes multiple measures of the arguments of the effectiveness function. Different measures would be appropriate depending on the availability of relevant data, the stage in the training-device development cycle, and the goals of the analysis. When multiple measures of the effectiveness factors are available, they are weighted using multiattribute utility assessment methods (MAUM), and the overall summary value for the factor scores is a weighted sum of the individual measures.

The overall strategy for determining the cost-effectiveness of a training-device design is to compare the effectiveness of the training device to the ratio of the hourly operating cost of the training device to that of the weapon system. This ratio is termed the operating cost ratio (OCR). This comparison is straightforward when effectiveness is measured by a transfer effectiveness ratio (TER; Roscoe, 1971). In this case the cost effectiveness is optimized by minimizing the ratio, OCR/TER.

Since the TER is only one of many possible effectiveness measures considered by TECIT, the straightforward comparison of TER and OCR is appropriate only for a limited number of situations. TECIT includes several decision rules that address considerations other than transfer as measured by the TER. However, the rules for these situations are incomplete, and the

model does not give adequate guidance when effectiveness measures other than TER are used.

In summary, TECIT provides a framework in which to incorporate many different cost and effectiveness measures. The overall effectiveness is determined by the weighted sum of these measures. TECIT does not specify how to compare the effectiveness of two or more training devices when different effectiveness measures are available for the two devices. TECIT provides ways to evaluate training-device cost effectiveness. However, these methods are only appropriate in limited range of situations.

Summary of Model Functions

The models described above perform several different functions. The relationship between model functions is shown in Table 1. As this table shows, the major functions served by the OSBATS model are media selection, training device design, training system evaluation, and cost evaluation.

The role of the OSBATS model in media selection is focussed on two specific functions: (a) selecting tasks that should be trained by a full-mission or part-mission simulator, and (b) assigning training on tasks to different training devices. Other procedures for media selection are much broader in that they consider a much wider range of training media. However, the two specific functions provided by the OSBATS model complement the functions provided by other media selection methods. For example, traditional methods could be used to identify the tasks that require device-based training. The OSBATS model would analyze these tasks further to determine the kind of training device that would best meet the requirements.

One of the OSBATS model's major functions is to aid training-device design. The model includes two modules that specifically address this problem. These modeling tools specify the instructional features and levels of fidelity that are best suited to the training requirements. The OSBATS model is the only one of the models reviewed that specifically addressed the device-design process.

The OSBATS model evaluates training devices as a component of the training system, unlike TECIT (Goldberg, 1988), which evaluates training devices as an individual entity. In this respect, OSBATS shares the characteristics of DEFT and CEMATD. However, as noted in the preceding discussion, the methods used by OSBATS differ considerably from those used by both of the other models.

Table 1

Comparison of Functions of Optimization Models

Function	TECEP	ISD	TASCS	CTEA	TDDSS	TSORT	CEMATD	DEFT
TECIT	OSBATS							
Select Tasks for Training		x		x	x	x		
Task Sequencing		x	x	x	x			
Media Selection	x	x	x	x	x			x
POI Development		x	x	x	x			
Training Device Design							x	
Training Device Evaluation						x	x	x
Training System Evaluation		x	x	x	x	x		x
Cost Evaluation	x			x	x	x	x	x

Training-Device Fidelity

The question, "How much fidelity is enough?" has been posed since the inception of training devices and simulators; it has been discussed in numerous reports and articles (e.g. Hays, 1980; Kinkade & Wheaton, 1972). Since both training-device cost and training effectiveness vary as a function of fidelity, a useful model of simulation must predict the relationships among training costs, training effectiveness, and device fidelity. This section discusses topics that are considered particularly germane in defining the relationships that are critical to a training-device cost-effectiveness model.

The first part of this section is devoted to discussing a definition of fidelity that we have conceptualized to structure our thinking about the design and necessary capabilities of a simulation model. In the second part of the section, we review research designed to assess the transfer of training from a flight simulator to the parent equipment.

Training-Device Fidelity

Attempts to formulate a suitable definition of training-device fidelity commenced more than 30 years ago. (See Gagne, 1954, Miller, 1954, and Adams, 1957 for early definitions of fidelity.) Since that time, fidelity has been conceptualized and defined in numerous and sometimes conflicting ways. Several recent reports (Hays, 1980; Ryan-Jones, 1984; Semple, Hennessy, Sanders, Cross, Berth, and McCauley, 1981) identify and discuss the various definitions of fidelity that have emerged during the last 30 years; all acknowledge that there is a lack of consensus about how best to define simulator fidelity.

Our review of the various definitions of training-device fidelity failed to reveal a definition that we considered to be entirely suitable for the purposes of this project, so we found it necessary to formulate yet another definition of fidelity. However, in formulating our definition of training-device fidelity, we have incorporated many of the fundamental ideas and observations that were originated by others. The ideas that have had the greatest influence on our conceptualization of fidelity are summarized below.

1. Central to our definition of fidelity is the concept, originated by Baum and his associates, that fidelity must be defined in terms of domain of interest (X), a referent (Y), and a metric (Z) (Baum, Smith, Hirshfield, Klein, Swesey, and Hays, 1982). Hence, a definition of fidelity must be of the form: fidelity of "X" relative to "Y" as assessed by the metric "Z."

2. The general domain of interest for the present study is training devices. However, because the fidelity of different components of a training device can vary independently, a clear understanding of the capabilities and limitations of a training device requires that the fidelity of individual device components be assessed individually as well as collectively.
3. The referent against which a simulator attribute normally is compared is the corresponding attribute of the equipment being simulated -- taking into account the mission, the full range of tasks that must be performed to accomplish the mission, the full range of environmental conditions in which the crew must be capable of performing the tasks, and, most importantly, the specific training objectives of the simulator. It is conceivable that a component of a high-fidelity simulator could serve as a referent in assessing the fidelity of the corresponding component of another, lower fidelity, training device. However, such a comparison is meaningful only if the training effectiveness of the referent simulator has been firmly established.
4. The primary metric of training-device fidelity is transfer of training. Because transfer of training can be measured only after a device component has been fabricated, and because transfer of training studies are extremely costly, numerous secondary metrics have been proposed. These secondary metrics are useful only to the extent that they are reliable and valid predictors of transfer of training.
5. Training device fidelity varies along two independent dimensions: realism and comprehensiveness. These dimensions are defined and discussed below.

Our definition of training device fidelity is characterized in Table 2 and is discussed below. The definition considers (a) the dimensions of fidelity, (b) a taxonomy of training-device attributes, and (c) a taxonomy of metrics used to assess fidelity.

Dimensions of Fidelity

Two dimensions, realism and comprehensiveness, are used to characterize both the fidelity of a training device and the differences between the fidelity of alternate training-device designs. Each of the two dimensions is further subdivided into sets of attributes.

Realism. The first dimension, realism, refers to the measured similarity between the training-device attributes and the corresponding attributes of the actual equipment. As conceptualized here, realism encompasses three classes of

Table 2

Characterization of Training-Device Fidelity

Dimensions of Fidelity		Taxonomy of Training Device Attributes						Taxonomy of Fidelity Metrics									
		Dimensions			Physical Measures			In-Device Responses			Analytic Measures		Training Transfer				
R	E	A	L	I	S	H	Static Display and Control	Layout of Instruments and Controls			Instrument and Control Design		Instrument				
							Dynamic Response Range	Motion System(s)			Motion System(s)		External Display				
C		O		P		R		Instrument			Instrument		Instrument				
O		O		H		E		External Display			External Display		External Display				
N		N		N		N		Control			Control		Control				
S		I		V		E		Performance Envelope			Performance Envelope		Performance Envelope				
T		E		E		E		Motion System(s)			Motion System(s)		Motion System(s)				
E		E		E		E		Instrument Readings			Instrument Readings		Instrument Readings				
S		S		S		S		Control Input/Feedback			Control Input/Feedback		Control Input/Feedback				
I		V		E		E		External Displays			External Displays		External Displays				
V		E		E		E		Individual			Individual		Individual				
E		E		E		E		Crew			Crew		Crew				
E		E		E		E		Team			Team		Team				
E		E		E		E		Combined Arms			Combined Arms		Combined Arms				
S		I		V		E		Equipment Malfunctions			Equipment Malfunctions		Equipment Malfunctions				
T		E		E		E		Degraded Visibility			Degraded Visibility		Degraded Visibility				
E		E		E		E		Adverse Weather			Adverse Weather		Adverse Weather				
S		S		S		S		Physical Stress			Physical Stress		Physical Stress				
I		V		E		E		Other Stress			Other Stress		Other Stress				
V		E		E		E		Varied Topography			Varied Topography		Varied Topography				
E		E		E		E		Varied Targets/Threat			Varied Targets/Threat		Varied Targets/Threat				
S		S		S		S		Visual			Visual		Visual				
I		V		E		E		Auditory			Auditory		Auditory				
V		E		E		E		Proprioceptive			Proprioceptive		Proprioceptive				
E		E		E		E		Kinesthetic			Kinesthetic		Kinesthetic				

attributes: (a) the configuration of the static displays and controls, (b) the dynamic response of all non-static components, and (c) the sensory stimuli generated by the training device. The realism of the sensory stimuli generated by the training device is related to both the realism of the static displays and controls and the realism of the dynamic response; however, the relationship is not perfect. For example, the dynamic response of a computer-generated extra-cockpit display may be highly realistic and yet, because of inadequate resolution or brightness, may fail to provide the visual stimuli required to perform a task.

Comprehensiveness. The second dimension, comprehensiveness, refers to the range of a device's potential training applications. As defined here, the comprehensiveness of a training device is characterized in terms of the following attributes: (a) dynamic response range, (b) the range of operational tasks that can be performed in the simulator, (c) the range of operational conditions that can be simulated, and (d) the range of sensory stimuli generated by the training device. The referent for evaluating a device's comprehensiveness is the actual equipment.

The relationship between realism and comprehensiveness. In principle, the two dimensions of fidelity should be treated independently. In practice, however, it makes little sense to assess a training device's comprehensiveness without taking into account its realism. It would be incorrect to describe a training device as being highly comprehensive if the realism of its components is so low that effective training on many relevant tasks is not possible. It seems more meaningful to describe comprehensiveness in terms of the range of simulated tasks, conditions, etc. for which realism is "adequate." The problem in implementing this sequential assessment of comprehensiveness, obviously, is defining the methods and metrics to be used to determine whether or not realism is "adequate."

The names used here to describe the two dimensions of fidelity are the same as those used by Jones, Hennessy, and Deutsch (1985), but the meaning of the names is somewhat different. Jones and his colleagues use the term "realism" to refer only to the "physical representation" of a simulator, and they use the term "comprehensiveness" to refer to "the degree of completeness and accuracy of representation of all functions, environmental characteristics, situational factors, and external events that are present in the target system or affect its function" (Jones, et al., 1985, p. 6). It appears to us that Jones and his colleagues use the term realism to refer only to the similarity between the displays and controls (static) and the corresponding simulator displays and controls (static) and use the term comprehensiveness to refer to both (a) the similarity between the dynamic response of the actual equipment and the

dynamic response of the simulator, and (b) the degree of completeness of the simulator's functional capability relative to that of the actual equipment. If our interpretation is correct, the dimensions are confounded in the sense that comprehensiveness encompasses both an element of realism and an element of completeness. The purposes of the present effort are best served by a definition that makes a clear distinction between realism and comprehensiveness for both the static and dynamic attributes of a simulator.

Relationship to the OSBATS model. The OSBATS Fidelity Optimization Module is organized around a set of dimensions that reflect training-device attributes that can vary in their sophistication, and consequently vary in their cost and effectiveness. The training device attributes addressed in the OSBATS model can affect both realism and comprehensiveness. Attributes that can enhance realism include such features as visual resolution, field of view, and platform motion. Attributes that can enhance comprehensiveness include such features as special training conditions, and visual or auditory special effects.

The two dimensions of training-device fidelity described here are reflected in two procedures that the OSBATS model uses to calculate the benefit of training-device attributes. First, realism requirements are evaluated on a task-by-task basis; the task requirements are compared to the capabilities offered by available levels of training-device attributes. Second, comprehensiveness is evaluated by aggregating effectiveness measures over tasks. This procedure ensures the training effectiveness measure obtained reflects the need to provide the range of conditions required to meet the training requirements.

Taxonomy of Training-Device Attributes

The taxonomy of training-device attributes provides a mechanism by which we can generate the fidelity alternatives needed for a specific application of the OSBATS model. The taxonomy of training-device attributes listed in Table 2 should be treated as preliminary; it seems probable that additional consideration will lead to modifications and refinements. Nevertheless, the present taxonomy is adequate to reflect our thoughts about the type of attributes that must be considered when assessing training device fidelity for each of the two dimensions defined above. Eight categories of training device attributes are described below.

Static display and control realism. The realism of a static displays is assessed with respect to the similarity between (a) the dimensions of the actual equipment and training device or simulator stations, (b) the layout of instruments and controls in the actual equipment and simulator, and (c) the design of the

instruments and controls in the actual equipment and in the simulator. As defined here, static display and control realism does not encompass the completeness of the instrument and control configuration; it refers only to the realism of the instruments and controls present in a particular simulator.

Dynamic response realism. The term "dynamic response fidelity" has often been used to refer to the fidelity of a simulator's software and hardware components, such as (a) the aerodynamic equations of motion, (b) the algorithms and hardware that drive the simulator's motion system(s), and (c) the algorithms that drive the image to the student. The term is used here in a similar but not identical manner that makes the concept more general. In the present case, dynamic response is defined only in terms of the realism of the inputs to and outputs from dynamic system components. That is, it is the inputs and outputs that are realistic, not the hardware or software that produces them. Specifically, the realism of the dynamic response of a training device is assessed in terms of the realism with which:

1. The training device responds to control inputs and the realism of the control feedback the student receives from the controls,
2. The simulated equipment's dynamic state is reflected in the displays on the instruments,
3. The simulated equipment's dynamic state is reflected by the motion system(s),
4. The simulated equipment's, dynamic state is reflected by the external display,
5. The simulated equipment's dynamic state is reflected in the audio generation components, and
6. Environmental conditions (including threats) and forces are reflected in control feedback, the instruments, the motion system(s), the external displays, and the audio generation systems.

Realism of sensory stimuli. As was stated above, realism of the sensory stimuli generated by a simulator is highly related to both static display and control realism and dynamic response realism. For this reason, considerable thought was given to excluding sensory stimuli from the taxonomy of training device attributes. However, our deliberations revealed several instances in which high static display and control realism and high dynamic response realism do not necessarily ensure high realism of the sensory stimuli produced by the simulator. The example mentioned earlier dealt with the realism of a computer-generated extra-cockpit visual display. Contemporary

computer-generated displays have a high degree of dynamic response realism, and yet, the visual stimuli may or may not be adequate to provide effective training on a given task. Similar comments can be made about the auditory generation system--the system that generates the sound associated with wind, rotor RPM, and certain types of equipment malfunctions. It is possible that the auditory generator could have high dynamic response realism with respect to its temporal response (onset of the auditory signal, temporal frequency of simulated rotor flap, etc.), and still generate an audio signal that is so dissimilar from the corresponding audio stimuli present in the aircraft that training effectiveness is degraded significantly. For these reasons, we concluded that, in some instances, realism cannot be fully assessed without considering the realism of the sensory stimuli produced by the simulator.

Static display and control comprehensiveness. Static display and control comprehensiveness is assessed by comparing the instruments and controls present in the training device with (a) the instruments and controls present in the actual equipment, or (b) the instruments and controls needed to accomplish all the training requirements established for the training device. As we have defined the terms, the dimensions and the layout of the instruments and controls are considered in assessing static display and control realism but are not considered in assessing static display and control comprehensiveness.

Dynamic response range. Dynamic response range refers to the range over which a simulator's components are capable of responding in a sufficiently realistic manner. Therefore, although dynamic response range is a different dimension from dynamic response realism, the former cannot be assessed meaningfully without considering the latter. Clearly, the number and type of tasks that can be trained effectively in a simulator are greatly influenced by the dynamic range of its components. The taxonomy listed in Table 2 shows that the simulator attributes that must be considered in assessing dynamic response range include the following:

1. The simulator's performance envelope, specified in terms relevant to the specific weapon system. For example, for flight simulators, the performance envelope would be expressed in terms of the minimum and maximum altitude, forward rate/acceleration, vertical rate/acceleration, lateral rate/acceleration, turn rate, torque, etc;
2. The simulator's motion system(s), specified in terms of the number of degrees-of-freedom and, for each degree-of-freedom, the maximum frequencies/amplitudes/accelerations, and the wash-out rates (for platform motion systems);

3. The simulator's instrument readings, as specified by the range over which the instrument readings remain valid and respond without excessive lags;
4. The simulator's control input and feedback, as specified by the controls that are present and operational and, for each operational control, the range of control inputs that are possible, the range over which control inputs cause valid and timely changes to the equipment state variables, and the range over which the simulation system provides valid and timely control feedback; and
5. The external displays (direct view and sensor), as specified by such factors as the maximum changes in system state parameters that are possible without excessive image smearing, image aliasing, or update lags.

Operational task comprehensiveness. A training device's value is heavily dependent upon its operational task comprehensiveness--the range of tasks that can be trained in the simulator. The importance of this factor is reflected in the OSBATS model, which obtains an overall fidelity-level benefit by summing the benefit value for each task. As is true for the other measures of training device comprehensiveness discussed above, operational task comprehensiveness can be indexed to the range of tasks that can be performed in the actual equipment, the range of tasks implicit in the training device's training objectives, or both. Soldiers are not permitted to practice some tasks on actual equipment because of accident risk and other constraints, so it is possible, in theory at least, that the ratio of training device training tasks to actual equipment training tasks could exceed a value of one. An assessment of a training device's operational task comprehensiveness should include individual tasks, crew tasks, team tasks, and combined arms tasks.

Operational conditions comprehensiveness. Success on the battlefield and survival in both combat and training environments are largely determined by a soldier's ability to function effectively under adverse conditions, such as adverse weather, inadequate lighting, equipment malfunctions, high enemy threat, and so on. Training on actual equipment under most adverse conditions is limited or, in some cases, prohibited because of the high likelihood of accidents. Furthermore, soldiers must be capable of operating effectively in a wide range of topographic contexts (desert terrain, mountainous terrain, rolling hills, built-up areas, etc.). One of the potentially greatest benefits to be realized from training devices is to enable soldiers to train under the adverse conditions and in the different topographic contexts that may be encountered in combat. For these reasons, the range of conditions and topography that can be simulated is an important index of the potential training

benefits that can be realized from a training device. Table 2 lists the general classes of conditions that should be considered when assessing the operational conditions comprehensiveness of a flight simulator. Specific examples of the conditions included in each class are presented below:

1. Equipment malfunctions: engine failure/damage, failure/damage of electrical components, failure/damage of hydraulic components, etc.;
2. Degraded visibility: darkness (with and without night vision goggles or other night vision aids) clouds, haze, fog, smoke, rain, and snow;
3. Adverse weather (effects other than visibility): heavy winds, wing gusts, wind shear, temperature and humidity extremes, etc.;
4. Physical stress: heat (when wearing Mission Oriented Protective Posture (MOPP) gear), exposure to chemical agents, exposure to nuclear contamination, etc.;
5. Other stress: high workload, distractions, fear, etc.;
6. Varied topography: varied terrain relief, vegetation, hydrography, cultural features (type/density), etc.; and
7. Varied enemy targets/threats: type, density, and distribution of ground and air targets/threats.

Comprehensiveness of sensory stimuli. Comprehensiveness of sensory stimuli refers to the extent to which the training device provides the full range of stimuli that are (a) available in the actual equipment, or (b) required to accomplish the specific training objectives established for the training device. The comprehensiveness of the sensory stimuli provided in a training device is assessed in terms of the types of stimuli that are present and the range of conditions and equipment states over which the stimuli remain sufficiently realistic.

Taxonomy of Fidelity Metrics

Central to virtually all definitions of fidelity, including the one proposed here, is the notion that fidelity refers to the degree of "correspondence" between the attributes of a training device and the corresponding attributes of the equipment being simulated. However, there is little agreement about the metrics that should be used to quantify degree of "correspondence." Vague metrics are implied by some definitions found in the literature. For example, the term "physical fidelity" implies that physical metrics are to be used to quantify "correspondence"; the term "perceptual fidelity" implies that

measures of human perception are to be used to quantify "correspondence." Specific examples of metrics implied by various definitions of fidelity are listed below:

1. Perception (of realism) (Gagne, 1954);
2. Physical, functional, environmental conditions (Miller, 1954);
3. Accuracy (Adams, 1957);
4. Contextual cues (Parker and Downs, 1961);
5. Looks (like), sounds (like), functions (like), and feelings/attitudes (toward aircraft/simulator) (Smode, Gruber, and Ely, 1963);
6. Missing, distorted, or misleading cues (Mudd, 1968);
7. Appearance and control feel, sensory stimulation, and perceived duplication (Kinkade and Wheaton, 1972);
8. Perception (of reality) and illusion (of reality) (Wood, 1977);
9. Layout, feel, stimuli, and responses (Condon, Ames, Hennessy, Shriver, and Seeman, 1979);
10. Behavioral and information-processing demands (Freda, 1979);
11. Correctness of psychomotor and cognitive control strategies (Heffley, Clement, Ringland, Jewell, Jex, McRuer, and Carter, 1981);
12. Type and consequences of errors (Heffley, et al., 1981); and
13. Effectiveness of learning and practice on specific tasks (Semple, Hennessy, Sanders, Cross, Beith, and McCauley, 1981).

Listing implied metrics entirely out of context, as has been done here, is clearly unfair to the various authors cited; all would undoubtedly argue that their definitions of fidelity were formulated to make a point about the factors that should be considered in assessing fidelity, and that the implied metrics do not represent their final thoughts about precisely what should be measured. Nevertheless, the above listing serves to illustrate that previous definitions of fidelity reflect diverse, and in most cases, very vague notions about the metrics that are to be used to quantify fidelity.

Table 2 shows a gross taxonomy of fidelity metrics. The cells that are marked with an "X" indicate the metrics that have been used or, in theory, could be used to assess the fidelity of the corresponding simulator component. As was stated earlier, it is our view that transfer of training should be treated as the primary metric of training device fidelity. That is, training transfer is the ultimate "proof of the pudding." High fidelity or low fidelity, as measured by other metrics, has meaning only to the extent that the measured level of fidelity is related to the amount of training transfer.

Other metrics, referred to here as secondary, are not unimportant; indeed, they serve at least three important purposes. First, secondary metrics are all there is to work with when acceptable training transfer data simply are not available at the time that important simulator design decisions must be made. For instance, when the aircraft simulators now being fielded by the Army were designed, the training transfer data available to support decisions about simulator fidelity requirements were (and still are) woefully inadequate. The authorities in charge apparently decided that the research needed to compile the requisite training transfer data would be too costly and too time consuming. So, the personnel responsible for evaluating the simulator design specifications had no alternative other than to employ secondary metrics to judge whether or not the proposed design would yield adequate fidelity. Second, even when training transfer data are available, secondary metrics may yield diagnostic information that is of great value in identifying beneficial design modifications and developing optimal training methods and procedures. And third, as additional data are accumulated and the relationship between training transfer and secondary metrics becomes better understood, it seems probable that models can be developed that provide the capability to accurately predict the degree of training transfer from some weighted combination of secondary metrics. Such a model, which would reduce the need for costly and time-consuming transfer-of-training research, would be an enormously valuable asset to the training community.

To complete our characterization of fidelity, it will be necessary to compile, for each metric class, a complete inventory of the specific measures that are needed to quantify realism and comprehensiveness for each attribute of a flight simulator. Although such a compilation is beyond the scope of this preliminary review, the following paragraphs present examples of specific measures that fall within each of five metric classes.

Physical measures. The Defense Science Board has stated that greater emphasis needs to be placed on the development of low-cost simulators that can be produced in far greater numbers than is economically feasible for the extremely costly full-mission simulators now being fielded (U.S. Department of

Defense, Defense Science Board, 1982). The call for greater emphasis on low-cost simulators is based partly on the belief that substantial cost reduction can be realized through a better use of technology, and partly on the belief that effective training can be accomplished with training devices whose physical attributes differ substantially from the corresponding attributes of the equipment being simulated. Support for the latter belief is provided by studies that have demonstrated that effective training transfer on some tasks can be achieved with training devices whose physical attributes are quite different from the corresponding attributes of the actual equipment. For example, it has been shown that procedures training in a photographic mock-up of a cockpit produced as much transfer as a high fidelity simulator (Dougherty, Houston, and Nicklas, 1957; Prophet and Boyd, 1970). Similarly, a high-percent transfer on traffic pattern flight and stall recoveries has resulted from training in a simulator whose visual system consisted of a stationary picture of the ground and horizon line and a line drawn on a blackboard to depict the aircraft's flight path (Flexman, Roscoe, Williams, and Williges, 1972).

Studies such as the ones cited above establish the fact that effective training on some tasks can be accomplished with training devices whose physical attributes depart dramatically from the physical attributes of the equipment being simulated. However, it would be both erroneous and misleading to assume that there is not a powerful relationship between a simulator's training effectiveness and its physical characteristics. Logic alone is sufficient to conclude that, as the physical characteristics of a simulator continue to depart from the physical characteristics of the actual equipment, a point will eventually be reached at which training transfer will decrease with further departures from physical correspondence.

Central to our views about fidelity assessment is the strong conviction that it is not possible to conduct meaningful analytic or empirical research on training device fidelity without using physical metrics to quantify the manner and degree to which a training device's attributes depart from the corresponding attributes of the actual equipment. It is the physical attributes that must be manipulated in order to vary fidelity, it is physical attributes that must be considered in estimating a training device's cost, and it is physical attributes that must be considered when developing training device design specifications. In short, metrics measuring the physical aspects are a necessary common denominator for designing fidelity research, evaluating the cost effectiveness of a training device, and translating fidelity research findings into training device design requirements. For most training device components, little attention has been given to the identification of (a) the specific design parameters that can be manipulated to vary

departure from complete realism and/or comprehensiveness, or (b) the specific metrics needed to quantify the degree to which each parameter departs from complete realism and/or comprehensiveness.

Although we have not made a concerted effort to develop a comprehensive inventory of physical metrics, we have given the matter enough thought to realize that such an effort will not be easy. The development of a metric with which to scale every parameter of every simulator component would be enormously difficult and time consuming. One way to pare down the job to realistic proportions is to first eliminate from consideration simulator components for which departures from realism or comprehensiveness would yield no significant cost savings. In other words, if an equipment component cannot be duplicated in the simulator at an acceptable cost, it makes little sense to expend resources to develop metrics and conduct the research needed to quantify departures from realism/comprehensiveness and the effect of such departures. For the remaining components, it will be necessary to identify specific parameters for which departure from realism and/or comprehensiveness is possible and promises non-trivial cost savings, and, for each parameter, to develop physical metrics that serve to quantify the degree of departure from realism/comprehensiveness.

As was suggested earlier, the derivation of parameters and physical metrics for some simulator attributes is certain to be a difficult task. The derivation of the parameters and metrics needed to quantify the scene content and scene-element design of a computer-generated, external display is certain to be among the most difficult tasks. The only metric that we know of that has been used to quantify a computer-generated scene is the number of basic elements (lines, polygons, bi-cubic patches, etc.) that are required to construct a scene or an object within the scene. Although this metric is useful for quantifying the proportion of a computer's capacity that is used to construct different scenes and objects, it appears to have little value in quantifying the departure of a computer-generated scene/object from its real-world counterpart.

Ratings. Ratings by subject matter experts have frequently been used in an attempt to quantify training device fidelity. In the most common case, aviators with considerable experience in the aircraft are required to fly selected tasks or missions in the simulator and are asked to make judgments about the realism of one or more simulator attributes. The judgments may be expressed informally during a debriefing session, or more formally through the use of rating scales specifically designed for this purpose. The use of aviator ratings as a metric for simulator realism has been roundly criticized by Adams (1979). His main criticism is aimed at the underlying assumption that there is a high positive correlation between amount of training

transfer and rated realism. He also questions the reliability of aviator rating data, citing research indicating that (a) aviator ratings of simulator realism are confounded with the aviators' experience in the aircraft, their experience in the simulator, and their individual skill deficiencies; and (b) aviator ratings of the realism of one simulator attribute are influenced by the degree of realism of other simulator attributes.

Although Adams' (1979) criticisms are valid, we believe that the problems he identifies reflect methodological errors and errors of interpretation rather than an inherent limitation of soldier rating data. With specially trained soldiers and with methods that offset the biases due to rapid accommodation to the simulator, it seems likely that the soldier ratings could serve as a highly useful metric of simulator realism, especially the realism of a simulator's dynamic response characteristics. The importance of special training is emphasized by Woomer and Carico (1977), who point out that a trend is underway in the Air Force to use specially trained engineering test aviators to assess the realism of the flight characteristics of simulators.

The Army is committed to the strategy of fielding training systems for new weapon systems at essentially the same time that the new weapon system is fielded. Although there are many good reasons to avoid long delays between weapon system delivery and training system delivery, the Army's current procurement strategy requires that many critical decisions about training device design be made before soldiers have an opportunity to acquire the weapon system experience needed to rate training device realism. So, the utility of using soldier ratings as a metric of fidelity depends upon the extent to which ratings of existing simulators are useful for (a) identifying ways to improve the training device being rated, and (b) predicting fidelity requirements for future training devices.

In-simulator responses. The assumption underlying the class of metrics referred to here as "in-simulator responses" is that useful information about training device fidelity can be gained from comparing soldiers' responses in the device with either (a) responses in actual equipment under comparable conditions, or (b) accepted performance standards. Listed below are examples of metrics that fall into this general class:

1. Peak performance level: a comparison of the highest level of performance achievable in the training device with (a) the highest level of performance achievable in the actual equipment, or (b) established performance standards;
2. Response strategies: a comparison of the cognitive and motor response strategies employed in the training device with those employed in the actual equipment under comparable conditions;

3. Errors: a comparison of the type, frequency, and consequences of cognitive and motor errors committed in the training device with those committed in the actual equipment;
4. Accuracy of absolute judgments: a comparison of the accuracy of absolute judgments of selected parameters made in the training device with (a) corresponding judgments made in actual equipment under comparable conditions, or (b) established performance standards;
5. Workload level: a comparison of the level of workload in the training device with the level of workload in actual equipment under comparable conditions;
6. Simulator sickness: a comparison of the incidence and symptoms of sickness experienced in the training device with that experienced in actual equipment under comparable conditions;
7. Eye movement patterns: a comparison of the patterns of eye movements (voluntary and involuntary) exhibited in the training device with (a) those exhibited in actual equipment or (b) those exhibited with different training device configurations (e.g., motion vs. no motion);
8. User acceptance: an assessment of user attitudes about the training utility of the training device, and an evaluation of the extent to which the device is being employed in an effective manner.

The use of in-simulator response metrics to assess simulator fidelity is appealing because the cost of compiling data on such metrics typically is far less than the cost of compiling data on many other metrics, especially transfer-of-training data. Furthermore, when responses in the training device are found to differ dramatically from responses in the actual equipment, it is logically appealing to conclude that the difference stems from non-trivial differences between the training device and the actual equipment. However, even a cursory examination is sufficient to reveal numerous questions, problems, and risks associated with the use of in-simulator responses as a metric of fidelity; a few examples are presented below.

Perhaps the most obvious and most critical question that can be asked about this class of metrics is: To what extent can training effectiveness be predicted from data on in-simulator responses and/or response differences? If effective training can be accomplished despite low correspondence between a training device and actual equipment measured by physical metrics, is it not possible that effective training can be accomplished despite large differences between responses in the simulator and

responses in the aircraft? We have been unable to locate any research specifically designed to determine the relationship between training effectiveness and any of the metrics listed above. So, for the time being, the credibility of such metrics must be assessed on logical grounds alone.

All of the metrics cited above, with the exception of user acceptance, require that responses in the training device be compared with responses in actual equipment, or, in some cases, performance standards. Although in-simulator responses usually can be measured with relative ease, measuring the corresponding responses in actual equipment may be a difficult problem. The problem may stem from the requirement of costly on-board instrumentation to measure responses in actual equipment. Metrics that suffer from this requirement include peak performance level, response errors, and response strategies. The problem also stems from the difficulty associated with ensuring that responses in the training device and responses in actual equipment are measured under comparable conditions. It may be difficult to define "comparable" conditions, and may be even more difficult to schedule the data collection effort at times and locations at which the desired conditions prevail. Regardless of the metric of interest, insuring comparable conditions of measurement is certain to be a difficult goal to achieve.

At least two metrics, peak performance level and accuracy of absolute judgments, are subject to serious confounding by artificial cues -- cues that may be present in a training device, but are never present in actual equipment. Ordinarily, such cues make the task in the device unrealistically easy. For example, a uniformly textured ground plane in a computer-generated, external display can make it unrealistically easy to perform some tasks on a flight simulator, e.g., nap-of-the-earth flight. At the same time, a uniformly textured ground plane can make it difficult to perform some other types of judgments, e.g., range estimation.

The above examples should not be taken as a complete indictment of the use of in-simulator responses as fidelity metrics. Rather, the examples were intended to illustrate some of the problems and risks associated with this class of metric.

Analytic measures. This class encompasses fidelity metrics, other than physical measures, that are derived analytically. There are at least two different sub-classes of analytic metrics. One sub-class includes metrics that serve to quantify the comprehensiveness of training device attributes. In their simplest form, metrics of comprehensiveness would consist of lists showing the range of tasks and conditions that can be trained in the training device relative to (a) the tasks and conditions specified in the device's training objectives, or (b) the full range of tasks and conditions specified in the weapon system's operational requirements. It should be a relatively

simple matter to compile lists that depict the comprehensiveness of the simulator attributes: displays and controls, operational tasks, and operational conditions. More thought and effort will be required to characterize the comprehensiveness of the attributes: dynamic response range and sensory stimuli. Although it is possible to derive a unitary metric that characterizes the comprehensiveness of the entire training device or the comprehensiveness of a specific device attribute, the purposes that unitary metrics would serve is not clear at this time.

Training transfer. Much has been said elsewhere in this review about the value of training transfer as a metric of simulator fidelity. Most of the training transfer research that has been conducted on simulators has employed the classical forward transfer paradigm that is designed to assess the extent to which training in the simulator transfers to the weapon system. Although the forward transfer paradigm is not without problems (see Adams, 1979; Blaiwes, Puig, and Regan, 1973; Matheny, 1974, 1975; and Mudd, 1968), it remains the most generally accepted paradigm yet developed. However, there are two other paradigms that may prove valuable for assessing training device fidelity.

One paradigm, referred to as a quasi-transfer paradigm, measures the extent to which training with one training device configuration transfers to another (usually higher fidelity) device configuration (see Lintern, Thomley, Nelson, and Roscoe, 1984; Sheppard, 1985). Quasi-transfer studies may prove to be a highly cost effective way to assess the relative fidelity of various device configurations. However, they are only appropriate if the training transfer of the high fidelity configuration has been firmly established.

Another paradigm that has potential value is the backward transfer paradigm. A "backward transfer study" is one that is designed to measure the degree to which actual performance skills transfer to a training device. Only highly experienced soldiers are used as subjects in a backward transfer study. The procedure is simple: an experienced soldier is placed in the training device and instructed to perform the task of interest without the benefit of practice. If the soldier is able to perform the task to criterion, a high degree of backward transfer is said to have occurred. The presence of backward transfer indicates that transfer from the training device to the actual equipment (forward transfer) is likely to be positive, but provides no information with which to estimate the magnitude of the forward transfer. The inability of experienced soldiers to perform a task to criterion in the training device must be taken as evidence of a problem with either the design or the functioning of the device. Hence, the absence of a high degree of backward transfer signals the need for further study of the training

device's characteristics to determine the reasons for the low backward transfer.

A variation of the backward transfer paradigm is to train the experienced soldiers in the simulator until their performance reaches an asymptotic level. This variation, of course, is appropriate only when there is a low degree of backward transfer. The nature of the learning curve in such cases provides useful diagnostic information. For instance, it must be concluded that the training device is either not providing the necessary cues or is incapable of processing control inputs correctly. Conversely, if the learning asymptotes at the criterion level after only a few practice trials, it can be concluded that the lack of high backward transfer is probably the result of minor differences between the stimuli and/or control responses of the training device and those of the actual equipment.

Implications for Modeling Effort

The definition of fidelity discussed above has a number of implications for developing a workable model for considering the tradeoffs among training device fidelity, training effectiveness, and cost. First, the model must be capable of accommodating a large number of attributes organized according to two dimensions of fidelity: realism and comprehensiveness. In dealing with comprehensiveness, the model must have an algorithm that prevents a simulator from being classified as highly comprehensive when the realism is so low that training transfer is improbable. That is, the algorithm must make comprehensiveness contingent upon "adequate" realism.

Second, because of the large number of training-device fidelity attributes, the model should have an algorithm that enables the user to identify and eliminate from further considerations device components that can be duplicated with minimal cost penalties. By duplication we mean the design of a simulator component whose physical properties and dynamic responses do not differ measurably from the corresponding component in the actual equipment. This capability will focus the analysis on the training-device attributes that have the greatest impact on cost.

Third, for device components that cannot be duplicated at a trivial cost, the model must be capable of accepting quantitative measures of the degree to which the physical attributes of the corresponding weapon system component, and must be capable of quantifying the relationship between cost and level of realism.

Use of these measures will allow the model to express device alternatives in a way that is consistent with the decisions that must be made in training-device design.

Fourth, the model must be capable of accepting inputs that serve to define:

1. A set of device training requirements--specified in terms of training tasks, environmental conditions, and criterion performance level for each task/condition;
2. The aptitude level and current level of relevant knowledge and skill of the trainee population;
3. The type and amount of training to be received on each task/condition;
4. The total cost of the training; and
5. The value of training outcomes (cost avoidance resulting from not having to train in the aircraft and the value of training that cannot be accomplished in the aircraft).

Fifth, the model must have an algorithm that predicts the training outcome quickly and reliably as the above conditions (fidelity, training requirements, student characteristics, and type/amount of training) are varied systematically. The training outcome should be specified in terms of training transfer or skill sustainment--whichever is appropriate for the application in question--and should include indices of training cost and value. This algorithm must be designed to accept and employ (in predicting training outcomes) training transfer data as well as data from research that has employed one or more secondary metrics of fidelity.

Transfer-of-Training Literature

The purpose of this subsection is to discuss the transfer-of-training literature as it bears upon the problem of developing models that have value in delineating the most cost-effective level of training-device fidelity for a given training application. The information contained in this subsection affects the rules that are used by the OSBATS model to derive fidelity requirements from task descriptions.

In order to provide a context for evaluating the utility of existing transfer-of-training literature, it is useful to consider the data required for the OSBATS model. First, data must be available to quantify, for each simulator design parameter, the relationship between the amount of training transfer and the level of realism and/or comprehensiveness. This relationship is central to the analyses performed by several OSBATS modules. Second, data are needed to quantify the manner in which training transfer is influenced by interactions among design parameters. Because of concerns for parsimony, the OSBATS

model assumes that design parameter interactions can be characterized by a simple multiplicative model. However, the form of the estimation function and the value of its parameters cannot be determined without data on the nature of such interactions. Finally, the model requires that the data base include the following types of cost and value data:

1. Definitions of training-device fidelity attributes and the levels of realism or comprehensiveness which they can attain,
2. Quantitative estimates of the realism or comprehensiveness associated with each level of each attribute,
3. Input parameters for an estimating function that relates the level of realism or comprehensiveness to cost, and
4. Input parameters that represent the maximum impact of each training-device fidelity attribute on transfer of training.

Other data used by the OSBATS model address the value of simulation-based training on tasks that cannot be trained on actual equipment.

Appendix A contains a synopsis of each of the flight simulator transfer of training studies that have been published in the literature since 1970, 26 studies in all. (Most of the training transfer studies published between 1970 and 1980 have been identified and reviewed by Semple and his associates (Semple, et al., 1981). Because of the focus of the initial OSBATS prototype, we have not included a review of the literature on procedures trainers and other part-task trainers. In addition, no attempt was made to review training transfer research on flight simulators developed solely for instrument flight training. The literature on instrument flight has been excluded for two reasons. First, the cost-effectiveness of instrument flight training simulators has been well established (for example, see: Caro, 1972, 1973; Povenmire and Roscoe, 1973; Roscoe, 1971). Second, the Army has no plans to procure additional flight simulators that are to be designed exclusively for instrument flight training; the capacity for instrument flight training is being designed into the full-mission flight simulators now being procured by the Army.

General observations about the set of transfer-of-training studies identified thus far are presented below. The reader is referred to Appendix A for specific information about individual studies.

Research Objectives

With few exceptions, the primary objective of the flight simulator training transfer research that has been conducted to

date has been to evaluate the training effectiveness of one simulator, configured in one way, and used for one training application (initial acquisition of basic flying skills, transitioning from one type aircraft to another, or skill sustainment of qualified aviators). So, for most studies, the primary independent variable investigated has been the presence or absence of simulator training prior to training to criterion in the aircraft. However, there are a few exceptions. The primary objective of three studies (Martin and Waag, 1978a, 1978b; Pohlman and Reed, 1978) was to determine whether the presence of platform motion contributes to the training effectiveness of the simulator. The presence or absence of motion was an independent variable in six other studies (Dohme and Millard, in preparation; Evans, Scott, and Pfeiffer, 1984; Gray and Fuller, 1977; Hagin, 1976; Jacobs and Roscoe, 1975; Ryan, Scott, and Browning, 1978), but determining the effect of motion was secondary to the primary objective of assessing training transfer from the simulator to the aircraft.

Only one study was located that investigated the relationship between training transfer and the design of the extra-cockpit display system (Thorpe, Varney, McFadden, LeMaster, and Short, 1978). Thorpe and his colleagues investigated the relative training effectiveness of a day/night color computer-image-generation system, a night-only, point-light-source computer-image-generation system, and a camera-modelboard system for training transition aviators to perform approaches and landings in the KC-135 aircraft. Although the results showed the two computer-generated display systems to be superior to the camera-modelboard system, the resulting data provide no specific information about either the factors that caused the difference in training transfer or the cost implications of the findings.

Other independent variables investigated in conjunction with the assessment of the flight simulator's training effectiveness include:

1. Aviator experience level (Brictson and Burger, 1976; Payne, et al., 1976),
2. Number of practice iterations during simulator training (Bickley, 1980),
3. Presence/absence of extra-cockpit visual display (Evans, et al., 1984),
4. Student aptitude (Gray and Fuller, 1977),
5. Presence/absence of g-seat motion (Hagin, 1976),
6. Supplemental visual cues (Lintern and Roscoe, 1978), and

7. Interspersion of simulator and aircraft training (Ryan, et al., 1978).

In summary, the training transfer studies conducted to date have been designed to evaluate a simulator rather than to define fidelity requirements. Not one training transfer study has been found that was designed for the express purpose of measuring amount of training transfer as a single dimension of simulator fidelity is varied systematically.

Aircraft Type and Stage of Training

Table 3 shows the distribution of transfer-of-training studies by aircraft type (military rotary wing, general aviation fixed wing, and military fixed wing) and stage of training (basic, transition, and continuation). It can be seen that the transfer-of-training studies that have been conducted to date clearly are not uniformly distributed across aircraft type and stage of training. Studies of fixed-wing aircraft simulators are far more numerous ($N = 20$) than studies of rotary-wing aircraft simulators ($N = 6$); furthermore, most of the fixed-wing studies have dealt with military ($N = 14$) rather than general aviation ($N = 4$) aircraft simulators.

All the studies conducted with general aviation simulators were designed to assess the simulator's utility for training basic flying skills to students with little or no prior flying experience. In contrast, the objective of most studies conducted with military aircraft simulators (both fixed and rotary wing) was to assess the simulator's effectiveness for transition training. Of the 26 training transfer studies located, only one (Holman, 1979) was specifically designed to assess a simulator's utility for continuation training. This observation is particularly significant in light of the fact that the Army plans to use about 85% of its flight simulators for continuation training.

Table 3. Distribution of Transfer-of-Training Studies by Aircraft Type and Stage of Training

Type Aircraft	Stage of Training		
	Basic	Transition	Continuation
Rotary Wing: Military	1	4	1
Fixed Wing:			
General Aviation	4	0	0
Fixed Wing: Military	2	14	0

Simulator Fidelity and Amount of Training Transfer

A primary reason for reviewing the transfer-of-training literature was to determine whether this body of literature contains data that could be employed to quantify the relationship between fidelity level and amount of training transfer. Our review of the literature has led us to conclude that the existing data are not adequate for this purpose. The considerations that have led to this conclusion are discussed below.

As was indicated above, our search failed to reveal a single study in which transfer of training was measured as the fidelity of a simulator component was varied systematically over several levels. Given that such studies have not been conducted, the next question is whether the results of studies conducted on different simulators can be synthesized in a way that enables one to draw valid inferences about the relationship between simulator fidelity and amount of transfer. It is indeed true that the simulators that have been used to conduct training transfer studies have varied widely in the fidelity of their components. However, these studies also have varied widely in such critical research design characteristics as the experience level of the aviators who served as subjects, the type of parent aircraft, the flying tasks investigated, the amount and type of training received in the simulator, and so on. The presence of these confounding variables makes it extremely risky to attribute differences in amount of transfer to differences in the fidelity of the simulators employed.

Because of the design of the training transfer studies conducted to date, it is risky to draw even very general conclusions from the data. For instance, consider the studies that have demonstrated positive transfer with very low-fidelity, extra-cockpit visual systems (e.g., Flexman, et al., 1972). Such studies demonstrate that some transferable skills on some basic tasks can be acquired in a simulator equipped with a very low-fidelity visual system. However, because only one level of fidelity was investigated, there is no way to determine whether (a) the transferable skills could have been acquired in a simulator with no visual system whatsoever, or (b) the amount of training transfer would have been far greater with a higher-fidelity visual system. Although it may be true that the fidelity level of contemporary simulators is excessive, it is clearly erroneous to assume that this claim has been established as fact by existing training transfer data.

The effect of motion on training transfer has received considerable attention and deserves special attention here. As was stated above, 9 of the 26 transfer-of-training studies listed in Appendix A investigated platform motion as an independent variable; 1 of the 9 studies also investigated g-seat motion as an independent variable. In every case, it was the presence or

absence of motion rather than the fidelity level of motion that was investigated. Not one study was found for which the presence of motion cues enhanced training transfer. Although these findings constitute sufficient justification for questioning the cost-effectiveness of platform motion on Army flight simulators, the findings are not sufficiently conclusive to justify the elimination of motion systems from existing and future flight simulators. Listed below are some of the reasons we believe the current body of research findings does not justify definitive conclusions about the need for motion systems on helicopter flight simulators.

1. Only one of the studies on the effects of motion has been conducted in a rotary-wing aircraft simulator (Dohme and Millard, in preparation). There are many reasons to argue that motion cues may be more important in rotary-wing than in fixed-wing aircraft.
2. All of the studies that have investigated the effects of motion have used relatively inexperienced aviators as subjects and have focused on the early stages of skill acquisition. Some experienced Army Instructor Pilots have argued that motion interferes with the early acquisition of flying skills, but that motion benefits skill acquisition and sustainment for more experienced aviators.
3. The lack of evidence that motion systems enhance training transfer may be due to unacceptable large lags in the motion systems, problems in the drive algorithms, inadequate synchronization of the visual and motion systems, the use of insensitive performance measures, or some combination of these factors. In short, the research results may simply show that no motion is no worse than bad motion.
4. The training transfer research has investigated only tasks in which motion feedback is the direct result of pilot control inputs; no tasks were investigated for which simulator motion is a joint function of control inputs and disturbances outside the pilot-aircraft control loop.

For the above reasons, we consider it unwise to exclude motion systems from the contemplated modeling effort. In fact, we believe that the model should include not only platform motion systems, but various force-cuing systems as well. Of the force-cuing systems that have been developed, only the seat shaker, the g-seat, and the stick shaker promise to provide cues that may replace or augment the cues generated by a rotary-wing platform motion system.

Cost and Value Data

Flight simulators have always been viewed as an economy measure in flight training. The supposition has been that relatively inexpensive simulator training can be used to replace some (preferably large) fraction of relatively expensive aircraft training in the attainment of a set level of flight proficiency. At the outset of the Army's SFTS program, the Army used the principle of economy through simulator-for-aircraft substitution in flight training as the primary purpose and justification for its flight simulation program. Given this simple supposition, the methods for quantifying the cost-effectiveness of flight simulators are straightforward. Given the requisite data on training transfer, simulator training costs, and aircraft training costs, Roscoe's Cumulative Transfer Effectiveness Ratios (CTERs) can be plotted and the cost-effectiveness of simulator training can be determined as a function of amount of simulator training (Roscoe, 1980, pp. 182-203). Povenmire and Roscoe (1973), Bickley (1980), and Holman (1979) have conducted studies in which CTERs were used to evaluate the simulator's cost effectiveness. The latter two studies determined CTERs for each of the sample of training tasks/maneuvers and a CTER for the composite training.

However, the Army no longer views simulator training as merely a means for reducing the aircraft hours and munitions required for training. For both initial-level (i.e., Aviation Qualification Course [AQC]) and continuation training, the Army views simulator training as a means to augment rather than to replace training in the aircraft. Flight hours and munitions allotted for training have decreased to such an extent that further reductions are not considered possible, regardless of how effective contemporary flight simulators prove to be. Rather, flight simulators are presently viewed as a means for (a) increasing skills beyond the level that can be achieved with aircraft training alone, and (b) for providing training on tasks that cannot be performed in the aircraft because of safety considerations or other constraints. So, the critical question is: Given that "X" number of aircraft flying hours and "y" amount of munitions (for attack aircraft) will be expended in training an aviator, what is the most cost-effective way to employ flight simulators to augment aircraft training? The traditional methods for assessing cost- effectiveness are not fully suitable for addressing this question. The main problem stems from the requirement to establish a dollar value of the increment in skill that results from simulator training. The following are offered as examples of the types of training outcomes for which dollar values must be established in order to evaluate the cost-effectiveness of using flight simulators to augment a fixed amount of aircraft training.

1. It is assumed that aircraft training plus simulator training results in more highly skilled AQC graduates than aircraft training alone. Is the value of the increased skill level great enough to offset the added cost of the simulator training?
2. Except during institutional training, aviators are prohibited from practicing certain emergency procedures (autorotations, hydraulic failures, antitorque maneuvers, etc.) in the aircraft because of the high cost of the accidents that occur during sustainment training on these emergency tasks. Simulator training on such emergency procedures has the potential for saving lives and reducing the cost of aircraft damage. To what extent does simulator training increase the probability of executing a successful landing in the event of an emergency? Are the savings (lives and property damage) that result from simulator training on emergency tasks great enough to offset the training costs?
3. Low-time unit aviators must accumulate a considerable number of aircraft hours before they are considered qualified to assume Pilot in Command (PIC) responsibilities. Does simulator training decrease the elapsed time and the aircraft hours required to become proficient enough to assume PIC responsibilities? Are the time and aircraft-hour savings great enough to offset the cost of training?
4. At some locations, local prohibitions prevent or limit nap-of-the-earth (NOE) training, night training, and weapons training. To what extent can such training be accomplished in the simulator? Is the value of such training great enough to offset the cost?

Although many other examples could be presented, the above are sufficient to illustrate that cost-effectiveness assessment of simulators, when used to augment aircraft training, cannot be accomplished without establishing the value of a variety of training benefits other than aircraft hours saved. Our review of the literature failed to reveal any instances in which attempts have been made to assess the dollar values of such benefits.

Instructional Features

Advances in computer and training technologies have promoted the development of a variety of simulator features designed to aid the process of instruction. For the present section, we pose a question analogous to the one posed in the previous section: Which set of instructional features provides the most benefit for the least cost? The present section addresses this question in terms of three related issues: what examples of instructional features have been cited in the literature, what empirical research has been done on the subject, and what are the rules for selecting one instructional feature over another? These issues are addressed separately in the following subsections.

Instructional Features in the Literature

Several studies have attempted to identify and describe instructional features that are currently available in flight simulators (Logicon, Inc., 1985; Sample, Cotton, and Sullivan, 1981; Caro, Pohlman, and Isley, 1979; Hughes, 1979; and Isley and Miller, 1976). Although these features are discussed in the context of flight simulation, most of them are sufficiently general in function to apply to other sorts of simulators as well. Table 4 summarizes the instructional features cited in each of these sources.¹ The table is arranged such that features listed within the same row share a common function even though they may have different names. In all, 25 instructional features may be distinguished by function. The first 10 of these are fairly well agreed upon in that three of the four sources provided some reference to them. The remaining 15 features are more idiosyncratic in that they are cited in only one or two of the sources. Each of the 25 instructional features is briefly described below.

1. **Malfunction control.** The purpose of this instructional feature is to provide instruction on emergency procedures, one of the most important functions of a training simulator. This feature allows the instructor to insert simulated malfunctions within a training scenario. Malfunctions may be inserted manually or automatically. In the automatic mode, a malfunction may be pre-programmed to occur under certain conditions or after a pre-specified period of time.
2. **Freeze.** This feature refers to the capability to stop all or selected parts of the simulation for the purposes of training. Action may be frozen manually by the instructor or may be automatically invoked under certain conditions,

¹The list of instructional features cited in Hughes (1979) was taken from Isley and Miller (1976); consequently, these two reports are regarded as a single source.

Table 4
Instructional Features Cited in Four Sources from the Research Literature

Logicon, Inc. (1985)	Sample, Cotton & Sullivan (1981)	Caro, Pohlman, & Isley (1979)	Hughes (1979)/ Isley & Miller (1976)
.Malfunction Control	.Manual and Programmable Malfunction Control	.Malfunction Simulation .Automatic Malfunction Insertion .Automatic Malfunction Insertion Exercise Preparation	.Preprogrammed Malfunction Activation
.Freeze	.Freeze	.Manual Freeze .Automatic Freeze .Parameter Freeze	.Freeze
.Simulator Record/ Replay	.Record and Replay	.Record/Replay	.Maneuver Playback .Audio Performance Record/ Playback
.Automated Simulator Demonstration	.Automated Demonstrations	.Demonstration .Demonstration Preparation	.Automatic Demonstration
.Briefing Utilities	.Recorded Briefings		.Automatic Briefing
.Scenario Control	.Programmed Mission Scenarios		.Non-Adaptive Training Exercises
.Initial Conditions	.Manual and Programmable Initialization		.Exercise Setup/ Initialization
.IDS Display Control and Formatting	.Annunciator and Repeater Instruments	.Remote Display	
.Automated Performance Measurement	.Automated Performance Measurement		.Automated Performance Measurement
.Hardcopy/Printout		.Hardcopy	.Hardcopy/Printout
.Tutorial	.ATD-Mounted Audio/Visual Media		
.Remote Graphics Replay	.Graphic and Text Readouts of Controller Information		
.Reposition		.Store/Reset Current Conditions	

Table 4 (cont.).

Instructional Features Cited in Four Sources from the Research Literature

Logicon, Inc. (1985)	Sample, Cotton & Sullivan (1981)	Caro, Pohlman, & Isley (1979)	Hughes (1979)/ Isley & Miller (1976)
.Automated Adaptive Training			.Adaptive Training Exercises
.Automated Controllers			.Ground Controlled Approach (GCA)
.Automated Performance Alerts			.Automatic Performance Monitoring/Alerts
.Closed Circuit Television			.Video Performance Record/ Replay
.Data Storage and Analysis			
.Real-Time Simulation Variables Control			
.Procedures Monitoring			
	.Automated Cuing and Coaching		
	.Computer Controlled Adversaries		
	.Computer Managed Instruction		
			.Automated Checkride
			.Automatic Copilot

e.g., a simulated "crash" or "kill." The extent of the freeze may vary from a total system freeze, in which all aspects of the simulation are frozen, to a parameter freeze, in which only a selected aspect of the simulation is frozen. An example of the latter is a flight system freeze wherein the simulator ceases to "fly," but all other components continue to function. Freezing the flight system is often used to train procedural components of flight tasks. Note that Hughes (1979)/Isley and Miller (1976) refer to parameter freeze as "performance-oriented guided practice," which describes an application rather than the function of the feature.

3. Simulator record/replay. The purpose of this feature is to allow the instructor to record a student's actions and inputs during a simulated mission and to replay it afterward for his review. Typically, the replay is temporarily stored in computer memory and limited to the last five minutes of performance. The record/replay feature is most useful when students are learning a new and difficult skill or when detailed performance feedback is required.
4. Automated simulator demonstration. The purpose of this feature is to provide a model of desired performance by allowing the instructor to pre-record and replay a maneuver. Although instructors can use the previous feature (simulator record/replay) to create demonstrations, the automated simulator demonstration feature differs from the previous feature in that the demonstrations are permanently stored. Also, demonstrations are not limited to five-minute playback periods. As in the previous record/replay feature, the automated demonstration feature is most useful when the student is learning a new and difficult skill.
5. Briefing utilities. The pre-training briefing serves to prepare the student for a particular training objective. The briefing may include a review of the student's past performance or an audio/visual description of an upcoming exercise. The briefing utilities feature refers to the capability to present this information automatically. The information may be in alphanumeric or graphic form and be presented via cathode-ray tube (CRT) display, as described by Logicon, Inc., or via sound recordings synchronized with an automatic demonstration, as described by Semple, Cotton, and Sullivan (1981).
6. Scenario control. This feature provides the instructor with the ability to configure and control the simulator so that simulated events occur according to a specific training scenario. Training scenarios are highly structured and meaningful sequences of events, such as takeoff under normal conditions or particular bombing maneuvers. The purpose of

this feature is to relieve the instructor workload related to controlling a complex training exercise. Note that Hughes (1979)/Isley and Miller (1976) distinguish scenario control involving pre-specified training events from that which employs adaptive training algorithms (described below as "adaptive training exercises").

7. Initial conditions. Prior to the start of a training session, the initial values of a variety of environmental and vehicle dynamics parameters must be pre-set. With the initial conditions features, sets of parameters can be pre-selected and stored to pre-set these values rapidly. This feature may be subsumed under scenario control.
8. Instructor operating station (IOS) display/annunciator and repeater instruments. As described by Logicon, Inc. (1985) and Semple, Cotton, and Sullivan (1981), the function of this feature is to provide the instructor with a display of current student performance during a simulated mission via the IOS. This information may be in the form of alphanumeric/graphical information presented on a CRT or repeater instruments that replicate information from the simulator cockpit. However, Caro, et al. (1979) described a slightly different function for the remote display feature: to present information simultaneously to the student and to the instructor. The purpose of duplicate displays is to facilitate communication between the two.
9. Automated performance measurement. The function of this feature is to calculate quantitative measures of student performance. The purpose of such measures is to assess student progress and provide information for diagnosing student performance problems. Usually, this information is not used as direct feedback to the student but is instead interpreted by the instructor who uses the information in his evaluation of the student. Also, the information provided by the performance measurement system provides input into other instructional features.
10. Hardcopy/printout. This feature creates a permanent paper record of the performance measurement data described in the previous feature. The record can be used to debrief students or to monitor student performance for course evaluation purposes.
11. Tutorial. The function of this feature is to provide training for student or instructor on the capabilities and appropriate uses of the simulator. This feature is essential if simulator training is intended to be self-administered. Semple, Cotton, and Sullivan (1981) describe the training in terms of slide/tape presentations, whereas Logicon, Inc. (1985) discusses this feature in terms of computer-assisted

instruction. Logicon also discussed a "help" function of the tutorial feature, which is designed to provide on-line assistance to the student or instructor.

12. Remote graphics display/replay. This feature provides a graphic or symbolic display of student performance. As described by Semple, Cotton, and Sullivan (1981), the function of this feature is to provide the instructor with an awareness of the current situation. But as described by Logicon, Inc. (1985), this feature includes record and replay capabilities in this feature. Thus, the latter authors propose that this feature also functions to provide detailed, post-training performance feedback to the student.
13. Reposition. This feature permits the instructor to position the simulated aircraft at a point in space that is relevant to the training scenario. This is a basic instructional feature that facilitates practicing especially difficult or critical portions of training exercises. Without this feature, the student or instructor must "fly" the simulator to a particular point, thereby wasting valuable training time.
14. Automated adaptive training. Adaptive training is an instructional approach wherein the difficulty of an exercise is tailored to the skill level of the student. Training begins at a relatively simple level and increases in difficulty dependent upon student performance. This feature typically allows the instructor to pre-select the adaptive variables. The computer then automatically sets the values of those variables according to some instructional sequencing algorithm, which itself is based on the student's performance on the previous trial.
15. Automated controllers. This feature provides for the generation of controller information for the pilot. This feature may be fully automated, meaning that computer-based voice recognition is used to interpret simple requests from the student and voice synthesis is used to present appropriate responses. In the less automated version of this feature, the computer calculates this information and presents it to the instructor. The instructor, acting as ground control, provides appropriate information to the student.
16. Automated performance alerts. This feature provides for an auditory or visual alert to be presented the student or instructor whenever performance tolerances have been exceeded. The purpose is to enhance the performance monitoring capabilities of both student and instructor. Of course, this feature requires that some meaningful tolerances can be established for the performance.

17. Closed-circuit television. A closed-circuit television system is used to monitor and record student behavior in the cockpit. Its purpose is to observe student behavior while the student is in the simulator and to replay it for him during the debriefing.
18. Data storage and analysis. This feature functions to store, analyze, and retrieve archival data on individual students, groups of students, or the simulator itself. The storage and analysis of individual data can be used in the pretraining briefing (see briefing utilities above), and group data can be used by course managers to evaluate the course (see hardcopy/printout above).
19. Real-time simulation variables control. This feature allows the instructor to insert, remove, and otherwise alter simulation variables during training, i.e., while the simulator is in operation. The most effective application of this feature appears to be for informal (i.e., continuation) training.
20. Procedures monitoring. This feature allows the instructor to monitor student performance of normal and emergency procedures. In a sense, then, this feature is analogous to the IOS display feature in that the former keeps track of discrete responding, whereas the latter monitors continuous responding.
21. Automated cuing and coaching. Similar to automated performance alerts, the automated cuing and coaching feature is activated whenever performance tolerances are exceeded. However, instead of (or in addition to) a warning signal, this feature provides a coaching message, which tells the student to take some corrective action. This feature appears especially appropriate for self-administered training.
22. Computer-controlled adversaries. In order to conduct tactical training, some sort of simulation of adversary aircraft is required. Computer-controlled adversaries (or so-called "iron pilots") are computer models that allow the simulation of enemy aircraft. The computer adversaries may be under partial instructor control or completely automated. Automated adversaries can also be made to differ in difficulty and can be used in conjunction with an adaptive training strategy.
23. Computer-managed instruction. This feature permits many of the instructional management functions to be assumed by computer. For instance, the computer can keep track of what objectives have been met and make appropriate assignments for subsequent exercises. Although Semple, Cotton, and Sullivan

(1981) could cite no simulators with this feature, they argued for its potential value.

24. Automated checkride. A checkride is a performance evaluation on a predetermined series of flight maneuvers for which performance standards have been established. This feature allows the simulator to administer and score the checkride automatically. Automation promotes a high degree of checkride standardization that would be impossible with human evaluators.
25. Automatic copilot. This feature allows the computer to assume the functions of the copilot. The automatic copilot is used mostly when a copilot is not available for training.

Empirical Research on Instructional Features

There are two major differences between empirical research on instructional features and research on fidelity features. First, the concept of simulator instructional features is newer, the term having been in use for around 15 years. Consequently, there are fewer empirical studies devoted to the subject of instructional features. Second, criterion measures for research on fidelity features and instructional features are fundamentally different. The purpose of simulator fidelity features is to maximize skill transfer from simulator to aircraft. Thus, as stated in the previous section, the "primary" measure is transfer of training. In contrast, the purpose of instructional support features is to increase the efficiency or effectiveness of the simulator. Thus, the appropriate criterion for instructional features is performance on the simulator itself.

The following review is divided into two subsections. The first subsection provides a review of research on the effects of some of the previously cited instructional features. The second subsection examines some related issues: how often instructional features are actually used in simulator-based training and the factors that determine their frequency of use.

Effectiveness of Instructional Features

Cross and Gainer (1985) identified only three empirical studies, all performed by Hughes and his associates, that specifically addressed the effectiveness of instructional features. All three experiments were conducted on flight simulators for fixed-wing fighter and attack aircraft. The first study (Hughes, Hannan, and Jones, 1979) compared the training benefits of using the automated simulator demonstration and the record/replay instructional features to the benefits of receiving an extra training trial on a cloverleaf maneuver. The use of the record/replay instructional feature was shown to be more effective than the use of the automated simulator demonstration.

However, an extra training trial was shown to be more beneficial than the use of either of the two instructional features.

The second study (Hughes, Lintern, Wightman, and Brooks, 1981) examined the effects of using the freeze and reposition instructional features. They compared performance in three experimental conditions: (a) a freeze/reset condition where the simulation was automatically frozen when an error was detected and reset to the correct position; (b) a freeze/flyout condition where the simulation was frozen as in the previous condition, but the student was required to fly out from the frozen position; and (c) a control condition where the freeze feature was not used. Analysis of performance indicated no differences among any of the three experimental conditions.

In contrast to the two previous studies that showed no training benefit from instructional features, Bailey, Hughes, and Jones (1980) showed that the initial conditions instructional feature can provide significant training value. They compared the effects of two training conditions on performance of a 30-degree dive bomb maneuver. In the control condition, students learned the maneuver in traditional "whole task" fashion, i.e., they practiced the task from beginning to end. In the experimental group, the task was divided into sequential segments, and the students learned according to a "backward chaining" schedule. The initial conditions feature was used to start the student at different points in the maneuver. The student was initially started on the final segment of the task. Only after the student had learned the final segment to criterion was he started at the next-to-last segment. Previous segments were added in a similar manner until the student practiced the entire task. The results showed that the experimental group performed significantly better and reached criterion faster than the control group, who did not have benefit of the initial conditions manipulation.

In interpreting these results, one must avoid implicitly accepting the null hypothesis: It would be inappropriate to conclude from these experiments that the initial conditions instructional feature is effective and that the automated simulator demonstration, record/replay, freeze, and reposition instructional features are not. It is quite likely that the training efficiency of an instructional feature is largely dependent upon the manner in which it is employed. Thus, a more appropriate interpretation of these data emphasizes the positive results of Bailey, et al., 1980. These experimenters showed that the initial conditions instructional feature can provide significant training benefits if combined with an effective training technique such as backward chaining.

Use of Instructional Features

Most of the generalizations concerning instructional features are based on anecdotal reports from simulator users. Although this can be an important source of information, the anecdotal nature of these reports leads one to question their reliability and validity. Pozella (1983) upgraded the quality of this information by systematically examining the patterns of use and the perceived training value of instructional features. His method was to survey 134 Air Force instructor pilots who use simulators to train their students. As expected, he found that instructional features vary with respect to the frequency that they were used. For instance, reset and flight system freeze instructional features were rated as being used often, in contrast to automated simulator demonstration and record/replay features that were rarely used. In addition, the frequency of use ratings were positively correlated with other ratings that measured the amount of training instructors have received on the feature, the feature's ease of use, and the training value that the instructors perceived the feature had. Pozella concluded the following about the use of advanced instructional features (AIFs) in aircrew training devices (ATDs):

The results of this survey indicate that most AIFs are under-utilized. The reason for this appears obvious: instructors typically receive minimal training in AIF use and, consequently, are not familiar with the AIF-capability of their respective ATDs. As training increases, AIFs become easier to use, their training value becomes more apparent, and they are used more often. (Pozella, 1983, p. 56)

Another notable finding from the Pozella (1983) study was a difference in usage patterns between instructors in replacement training units and instructors in continuation training units. Replacement training units concentrated on procedural training, whereas continuation training focused more on the tactical aspects of flight. Consequently, instructor pilots in replacement units tended to rate features such as flight system freeze more highly than instructor pilots in continuation training units. Freezing the flight system allows the student to practice procedural skills in isolation from flying the aircraft. In contrast, instructors in the continuation training units rated the scenario control feature higher, because it allows the instructor to preprogram a complex tactical scenario. Overall, instructors in replacement training units used instructional features more often than those in continuation training units. This latter finding tends to support the commonly assumed notion that instructional features are more appropriate for initial-level training as opposed to more advanced training.

Specification of Instructional Features

A basic requirement of the model is that it must specify a set of optimal instructional features for a particular application. Other researchers (e.g., Caro, Pohlman, and Isley, 1979; Semple, Cotton, and Sullivan, 1981; and Logicon, Inc., 1985) have also perceived the need for simulator design guidance with respect to specifying such features. The Logicon report represents the current state of the art because of its chronological relationship to the other reports and because it draws upon much of the earlier work. The OSBATS model, in turn, draws upon the Logicon work, with several changes. The purpose of this section is to review this latest guide in detail to identify procedures for selecting instructional features. Of particular interest are objective procedures that are sufficiently well developed to be implemented within a training-optimization model.

A seemingly basic assumption of the Logicon procedure is that "...before any ATD [aircrew training device] is specified, a front-end training analysis must be accomplished to determine ATD capabilities to support training" (p. 74). The front-end analysis must specify (a) the skills and knowledges of the student population, and (b) the training objectives (tasks) including all relevant conditions and standards. The product of the training analysis is a training syllabus that organizes training objectives into meaningful training scenarios. The requirement to perform a complete front-end analysis seems to be unnecessarily burdensome for the training-device designer. Furthermore, after examining the Logicon procedures rather closely, it is apparent that this requirement is probably overstated. Only two training analysis products are necessary for specifying tasks: (a) a general description of the skill level of the student/user (as opposed to a complete inventory of skills and knowledges of the training population), and (b) a list of the tasks that are to be trained on the simulator. A detailed training syllabus is not required for the following procedures.

As described in the guidebook, the process of specifying instructional features can be conceived as consisting of three sequential stages. The first stage of the specification process is to select instructional features that are relevant to a particular training application. The selected features are then prioritized with respect to their potential benefits. Finally, the cost and implementation factors are considered in making the final specification of features. Each of these processes is described below.

Selection of Instructional Features

The first stage of the process is to select instructional features that are relevant to the application in question. The guidebook does not provide well developed procedures for this stage in the process; rather, advice may be more accurately described as "rules of thumb." However, examination of these rules revealed three factors that are prominent in the decision whether or not to select a factor: (a) instructor functions, (b) student skill level, and (c) task characteristics. Each of these factors is discussed below.

Instructor functions. A basic premise of the Logicon guidebook is that instructional features are designed to support the instructor in the training process. Thus, the selection of relevant instructional features is based on an analysis of the instructor's role in simulator-based training. The report identified eight commonly accepted instructor functions: instructor training, briefing, controlling, monitoring, instructing, evaluating, debriefing, and recording. For a particular application, each function should be considered separately in order to identify instructor needs. Then a determination should be made whether or not that function would be facilitated by the corresponding instructional features. The problem with this analysis is that it is difficult to envision applications where instructor needs differ in some systematic manner. The guidebook is not helpful in this regard. In order to make this selection factor more usable, the relationship between instructor needs and specific simulator situations must be explained more fully.

Skill level. Pozella (1983) found that use of instructional features varies as a function of the skill level of students. Accordingly, the second factor that should affect instructional feature selection is skill level. The most important distinction in skill level is that between novice level training (e.g., undergraduate pilot training) and advanced level training (e.g., continuation training). In general, most of the features are appropriate to either level. Exceptions include three features that appear to be designed especially for beginners, as opposed to more advanced students:

1. Freeze
2. Simulator record/replay
3. Automated Simulator Demonstration

On the other hand, one feature, real-time simulation variables control, is best suited for advanced and not for beginning students.

Task characteristics. The third factor in the feature selection process concerns certain characteristics of the to-be-

trained tasks². These characteristics fall into two groups. The first characteristic is the extent to which task performance is dependent on procedural skills. If a particular task has a significant procedural component, then the feature, procedures monitoring, applies. If emergency procedures in particular must be learned, then the feature, malfunction control, applies in addition to the former.

The second task characteristic is difficulty. "Difficulty" is defined with respect to the extent to which some form of part-task training is required to learn the task. Part-task training may be required for the following sources of task difficulty: (a) the task is exceedingly long, (b) a segment or portion of the task is especially difficult, or (c) the task is "saturated" in the time-sharing sense, i.e., the performer must execute multiple actions simultaneously. In order to accomplish part-task training, the following instructional features are required:

1. Initial Conditions
2. Reposition
3. Freeze

The remaining question for the feature selection process may be phrased as follows: How should these factors be combined in order to decide whether or not a feature is selected? The Logicon, Inc. (1985) guidebook states that a feature should be considered for selection if it supports either "...the instructional objectives or instructor task" (p. 74). This suggests that a simple "or" rule could be used to select a feature. That is, experts could "tag" features according to the factors discussed above. Then, selected features would be those that are associated with at least one or more tags.

Benefit Analysis

The next stage in the process is to prioritize instructional features that have been selected on the basis of the potential training benefits that instructional features may accrue. Five types of benefits are discussed in the report:

1. Frequency of identified need. Those features used more often should receive a higher priority. The data reported by

²It is interesting to note that Semple, Cotton, and Sullivan (1981) argued that there is no relationship between task characteristics and selection of instructional features. These authors maintained that the selection is only dependent upon the student's level of training. In emphasizing the importance of training level, these authors are probably guilty of overstating their case.

Pozella (1983) may be useful in this regard. However, a low rating on this dimension may be overruled if a particular feature provides the only way to train a critical task.

2. Instructor loading. Features that substantially reduce the instructor's workload should be given a higher priority. Again, this relates to the central premise of the Logicon guidebook.
3. Useability of the system. Related to instructor loading, this benefit concerns the amount of time the instructor spends controlling the simulator relative to the total time instructing the student. A feature requiring a considerable portion of the instructor's time and effort should be given a lower priority.
4. Training efficiency. Efficient instructional features are those that allow more instruction to be accomplished in a given period of time. "For example, a remote briefing utility console or a remote graphics replay console would allow pre-training or post-training functions to be carried during the 'hands on' training" (Logicon, Inc., p. 82).
5. Instructional feature interdependency requirements. Instructional features are not independent. For instance, the functions of a feature such as initial conditions may be subsumed by a more general feature such as scenario control. These dependencies "should be a consideration in the prioritization of the selected features."

In order to be implemented within the context of the optimization model, the benefit analysis could be performed in the following manner: Scales would be developed to allow experts to rate instructional features with respect to the five dimensions discussed above. MAUM methods could then be used to combine ratings and make the appropriate decision.

Cost and Implementation Considerations

The final stage in the specification process is to consider the costs of implementing the instructional features. This stage is saved for last so that costs do not drive the feature specification process. Clearly, precise cost data are not available for each particular feature. Instead, costs should be considered in terms of the general architectural components of the simulator. These components are presented from most to least important in terms of their impact on simulator costs.

Task modules database. Task modules database refers to the computer files that relate to specific task modules. Task modules are the components of complete ("chock-to-chock") training scenarios; they ideally correspond to specific training

objectives such as Perform Pretakeoff Procedures, Identify and Correct Hydraulics Malfunction, etc. These modules specify how, when, and under what conditions an instructional feature is to function. The cost of the task modules database is directly related to the number and the complexity of the tasks demanded by the training scenarios. The report notes the following important qualification: "From a system resources point of view, the task modules reduce the amount of data that requires monitoring to encompass only those events that are critical at a specific point in time" (Logicon, Inc., p. F-4). In addition, "software modularity" (see below) with respect to training objectives is a desired characteristic of the data base.

Software. Separate software "modules" should be developed for each instructional feature. Software modularity allows users to add features easily at later dates and is thus a desired characteristic. Software should also provide for editors to generate and edit the task modules database.

Computer system. Three cost factors related to the computer system impact instructional feature specification: (a) processing capacity, (b) main memory, and (c) mass storage. The latter two cost factors are particularly important.

"Contributions to the storage requirement by each [feature] should be estimated using the stated functional requirement and the expected utilization" (Logicon, Inc., p. F-13). Unfortunately, the "stated functional requirement" appears to be a conceptual entity, rather than a product of any analysis; thus, its relationship to memory storage requirements is not known. As an example, the report uses the Record/Replay feature: The memory requirements for this feature are "...large and need to accommodate the total number of minutes which are to be recorded..." (Logicon, Inc., p. F-13). This example begs questions such as "What is large?" and "What is the relationship between minutes and Kbytes of storage?"

Stations. Stations are defined as the "person-machine interface between all users of the [simulator] and the [simulator itself]...". Three factors pertain to this cost consideration: (a) types of devices located at the stations, (b) number of stations, and (c) location of the station. The costs of such devices are dropping, so the specification process should only be based on the most current cost data.

Simulation system interface. This consideration refers to the interface of the instructional features with the rest of the simulation system. If instructional features are "...added to an existing ATD, the simulation system interface involves risks and possible interference with the operation of the ATD. The risks can be minimized if the [instructional features are] designed into the initial procurement of an ATD" (Logicon, Inc., p. F-18).

Life-cycle cost. Relevant life-cycle-cost factors include the following items: (a) number and type of personnel needed to run the simulator; (b) the operational design of controls, displays, and procedures; and (c) instructor/operator training helps and documentation. For instance, the tutorial instructional feature may directly impact the last factor by lowering costs associated with instructor/operator training.

It is certainly clear that these cost considerations are important factors for designing a training simulator. However, except in a few cases, it is not clear how the considerations relate to the specification of individual instructional features. As currently stated, the cost factors are not sufficiently well developed to be included in the optimization model.

OSBATS Instructional-Feature Selection Procedures

The OSBATS procedure for instructional-selection draws on the Air Force procedures, but is different in several respects:

1. It has a somewhat more limited scope in the class of instructional features it considers. OSBATS only considers those instructional features that make training more efficient in terms of the improvement in performance that can occur with a fixed amount of training time. It is not concerned with instructional features that serve primarily a training management, performance recording, or data analysis function.
2. Task characteristics play a much more central role in the process than in the Air Force model.
3. The OSBATS model brings cost into play at an earlier stage in the analysis. By combining cost and benefit in a benefit/cost ratio, it produces a superior solution to those procedures that determine a priority according to benefit, and then select according to the priority until the budget has been met.

The general steps of the OSBATS procedure are described below.

1. Task matching. The task characteristics and skill level are compared for each task to determine which instructional features are appropriate for each task.
2. Benefit assessment. An overall benefit is determined by aggregating the task matches over task. Each task receives a different weight in the aggregation process. The overall benefit for each instructional feature is further multiplied by an instructional feature weight that represents the frequency of identified need.

3. Benefit/cost ordering. The aggregated instructional-feature benefit is divided by an assessed cost. The instructional feature priority is determined by the benefit/cost ratio.
4. Instructional feature selection. The instructional features are selected by choosing the features from the top of the benefit/cost priority list until the budget has been met.

The OSBATS procedure encompasses many of the features of the procedure developed for the Air Force (Logicon, Inc., 1985). However, some considerations are omitted from the OSBATS analysis, namely, instructor functions, instructor loading, instructional feature interdependency requirements, and detailed cost considerations. Some of these considerations were not incorporated into the OSBATS model because of the limited scope of the OSBATS analysis, and because items such as instructor function are not currently part of the training-device definition used by the OSBATS model.

Models of Human Learning, Retention, and Transfer

The ability to produce optimal training-device designs depends upon being able to predict the speed with which skills are acquired as a function of training-device design, the extent to which these skills will be retained over time, and the degree to which skills acquired on a training device will transfer to actual equipment under operational conditions.

Models of Skill Acquisition

Several early experimental psychologists attempted to characterize the regularity in the learning process, particularly the relationship between the amount of practice and performance (e.g., Thurstone, 1919; Snoddy, 1926). Within the last 25 years work has focused on the development of detailed mathematical models of learning of simple tasks (Atkinson, Bower, and Crothers, 1965; Restle and Greeno, 1970). In more recent work, attempts have been made to integrate models of learning with performance models in complex cognitive (Anderson, 1982; Card, Moran, and Newell, 1983) and procedural (Sticha, Edwards, and Patterson, 1984) tasks. This section briefly reviews the form of skill acquisition models and discusses the scope of their application.

A theory of skill acquisition must address two important processes: (a) processes by which information to be learned is encoded in memory, and (b) processes by which the information is later retrieved. Most of the theories discussed here address both acquisition and retrieval, although greater emphasis may be placed on one or the other process.

The Power Law of Practice

The power law of practice represents an empirical relationship between performance and practice that applies to a wide variety of tasks and performance measures. This relationship states that the time required to perform a task varies as a power function of the amount of practice, or

$$T = BN^{-k}, \quad (1)$$

where T = the time required to perform the task;

B = the initial level of performance;

N = the number of trials; and

k = the learning rate.

Newell and Rosenbloom (1981) reviewed a number of studies to illustrate the variety of tasks for which the power law of practice holds. These tasks include psychomotor tasks, perceptual tasks, memory tasks, elementary decisions, complex procedures, and problem solving. The law holds for measures of

performance other than speed, such as accuracy measures (Stevens and Savin, 1962), although the evidence for the law is not as strong for these measures. In addition to showing the robustness of the power function, Newell and Rosenbloom demonstrated the empirical superiority of the power function over exponential and hyperbolic functions.

Attempts to develop a theory explaining this relationship have had limited success. Lewis (1979) suggested that performance improves according to a power function because a task is a combination of many components, each of which is improving exponentially. With proper, and rather restrictive, assumptions about the relative contribution of individual processes to overall performance, a power function relating speed and practice can be derived. Newell and Rosenbloom (1981) suggested that the power law is a result of chunking processes by which simple general rules for performance are combined into larger, more specific rules. The chunking theory also depends on restrictive environmental conditions to produce a power function, although the chunking model generally behaves similar to a power function. Neves and Anderson (1981) proposed a similar theory; they claim that improvements in performance speed are caused by a number of processes operating on the rules that produce skilled performance. According to this theory, learning processes translate knowledge from a general, declarative representation, to a procedural representation that incorporates specific knowledge of the skill being learned.

Suppes, Fletcher, and Zanotti (1976) were more successful in developing a general psychological theory that produces a power law of learning from five learning axioms. This theory is couched in a method for predicting a single student's progress through a curriculum. They used this method to control the amount of time the student spends on the curriculum and to achieve specific objectives in level of proficiency. Their five learning axioms were based on the following definitions:

$y(t)$ = position of student in the course,
 dy/dt = rate of progress through the course,
 $A(t)$ = cumulative amount of information introduced in the course up to time t ,
 dA/dt = rate of information introduction in the course, and
 $s(t)$ = student's rate of sampling information.

The axioms are:

1. A student's mean rate $s(t)$ of processing or sampling information is directly proportional to the rate of introduction of information and inversely proportional to the total amount of information introduced up to a time t : $s(t)$ is proportional to $(dA/dt)/A(t)$.

2. Upon introduction of a new piece of information, a student's new mean rate of processing information is decreased by an amount equal to the product of his current rate and the difference of his current rate and his asymptotic rate. For a small interval of time h :

$$s(t + h) = s(t) - [s(t) - s(\cdot)] s(t). \quad (2)$$

3. The probability of a new piece of information being introduced for a given time t is independent of t and the previous introduction of information.

4. The position of a student in a course is directly proportional to the total information introduced thus far in the course: $y(t)$ is proportional to $A(t)$.

5. The rate of progress of a student in a course is directly proportional to the rate of introduction of information in the course: dy/dt is proportional to dA/dt .

Suppes, et al. (1976) expressed dissatisfaction with the absence of a more fundamental characterization of the rate assumption in the second axiom. They suggested an alternative formulation: the decrease in rate of processing, upon introduction of a new piece of information, is quadratic. They derived a set of differential equations for $A(t)$ and $y(t)$. The derivation assumed that the student has some knowledge c initially (time $t = 0$), and the final equation is:

$$y(t) = bt^k + c, \quad (3)$$

where b = a scaling constant;
 c = initial performance; and
 k = learning rate or ability.

These parameters are estimated for each student and each task.

This model is nonlinear, but its parameters can be estimated using an algorithm developed by Golub and Pereyra (1972; cited in Fletcher, 1985). Suppes, et al. applied the model in an elementary mathematics curriculum. They estimated the constants of integration for each student and obtained a reasonable fit of the theory to data, as measured by mean standard error. One difficulty in application noted was the need for a detailed analysis of the curriculum to determine the strands, or components. On the positive side, the "concept of a student's mean stochastic trajectory is robust with respect to a variety of assumptions about the learning of individual items or component skills in a course" (Suppes, et al., 1976, p. 127).

It should be noted that despite the apparently wide support for the power law of practice, other forms of the learning

function have been proposed, and have received some support. In fact, as pointed out by Mazur and Hastie (1978), almost any shape of the learning function can be obtained under some conditions, including S-shaped curves, and even positively accelerating curves. They propose a hyperbolic form of the learning function; this form may be derived from the "accumulation model" of learning that was first proposed by Thurstone (1919). Spears (1985), on the other hand, has proposed a learning curve characterized by a logistic function, which produces an S-shaped learning curve. The logistic function provided a good fit to flight training data, although the data generally fell on the negatively accelerated part of the curve.

Threshold Models

One of the oldest of the mathematical models of learning was developed by Hull (1943, 1952) and Spence (1956). This class of models is named for its retrieval mechanism, which produces a response when association strength exceeds a threshold. Early mathematical learning theorists rejected the Hullian approach because of difficulties in estimating its parameters. Most of these problems have been resolved by the development of computer routines that search through a space of parameters to find those values that minimize a function specified by the user.

Hull's learning model postulated an exponential increase in strength to an asymptote:

$$H_k = M - (M - H_1)k^{N-1}, \quad (4)$$

where H_k = the strength of the response,
 H_1 = the initial value of the strength,
 M = the strength asymptote,
 k = the learning rate, and
 N = the number of learning trials.

It was assumed, in accord with Thurstone (1927), that strength is best described by a normally distributed, random variable with constant variance. The likelihood of a correct response on a trial is the probability that strength exceeds a response threshold. (An alternative formulation, proposed by Grice, 1968, postulates a constant strength and a variable threshold; the two formulations predict equivalent acquisition curves.) The threshold model has five parameters: H_1 , M , k , the standard deviation of the strength distribution, and response threshold.

Versions of the threshold model have formed the basis of a number of approaches to acquisition, retention, and retrieval. For example, Wickelgren and Norman (1966) and Norman (1966) applied the threshold model to a short-term-memory experiment. Their experiment primarily investigated the implications of the threshold model for retention and retrieval. More recently,

Anderson (1982) has included concepts from the threshold model as one of several mechanisms to control retrieval of procedural knowledge. Similar mechanisms have been proposed by Raaijmakers and Shiffrin (1981).

Sticha et al. (1984; Knerr and Sticha, 1985) applied a Hullian model to eight military procedures. Their study investigated acquisition of these skills by soldiers receiving One Station Unit Training (OSUT), and retention of the skills by both students in OSUT and soldiers in an operational unit. The Hullian model was tested by comparing it to both simpler and more complex alternative models. In all cases, the Hullian model predicted the acquisition data better than simpler models. The resulting models predicted overall speed and accuracy well, both for the data on which the model parameters were estimated, and for a second portion of the data that was used for validation. However, there was evidence that model parameters varied among both individuals and task elements.

The parameters of the models described above were all estimated from performance data. However, in any application of a learning model for training system design, performance data are not, in general, available. It is necessary, therefore, to develop methods to estimate parameters of learning models from data that can be uncovered through task analysis. To this effect, Sticha and Knerr (1984) attempted to relate parameters of the Hullian learning model to subject-matter-expert assessments of fourteen task characteristics that were hypothesized to affect the learning process, and to individual aptitudes as measured by the Armed Service Vocational Aptitude Battery (ASVAB). The results showed significant relationships between both task and individual variables and the parameters of the learning model. However, the relationships were not consistent across tasks, particularly for individual variables. The authors suggested that some of the inconsistency in the results may be due to the large number of model parameters compared to the total number of degrees of freedom in the data.

Linear Models

Linear-learning models are different from threshold models in that the strength of an association in a linear model is equivalent to the probability of a correct response. A number of early learning models were of this form (e.g., Bush and Mosteller, 1955; Estes, 1959), but this approach has not been used recently because of research indicating that other models give a better account of learning data. The simplest of the linear models asserts that the probability of making a correct response to a stimulus increases according to the following linear operator:

$$P_{n+1} = P_n + r (1 - P_n), \quad (5)$$

where P_n is the probability of a correct response on trial n , and r is the learning rate.

Equation 5 indicates that the probability of an error decreases at a constant rate at each learning trial. Thus the probability of a correct response as a function of trials is given by the equation:

$$P_n = 1 - (1 - P_1)(1 - r)^{n-1}. \quad (6)$$

Figure 2 shows an example of such a curve for $P_1 = 0.5$ and $r = 0.25$.

The characteristic of the linear model that distinguishes it from other models is the distribution of errors on trials before the last error. The linear-operator model predicts that the distribution of errors is independent of the occurrence of the last error. This property of the model is the reason that it does not make any predictions regarding sequencing of instruction. Thus, the probability of a correct response on trials before the last error is as described in equation 6 and illustrated in Figure 2. The all-or-none model described next yields the same formula as shown in equation 6, but gives a considerably different prediction regarding the distribution of errors.

All-or-None Models

In contrast to the models described previously, all-or-none models of learning represent learning as an association's transition between a small number of states. The simplest of these models, called the one-element model, has only two states: unlearned and learned. Figure 3 shows network and matrix representations of this model. An association is assumed to start in the unlearned state; the probability of a correct response in this state is a guessing probability, g . On each trial, there is a probability, c , that the association will move to the learned state. In the learned state, performance is perfect. This model assumes that once in the learned state, an association will remain there.

The probability of a correct response can be calculated for the one-element model as it was for the linear-operator model. For an error to be made on trial n , the association must fail to be learned on $n - 1$ trials, which occurs with probability $(1 - c)^{n-1}$. In addition, an incorrect guess must be made, which occurs with probability $1 - g$. The learning curve, which expresses the probability of a correct response as a function of trial, is given by:

$$P_n = 1 - (1 - g)(1 - c)^{n-1} \quad (7)$$

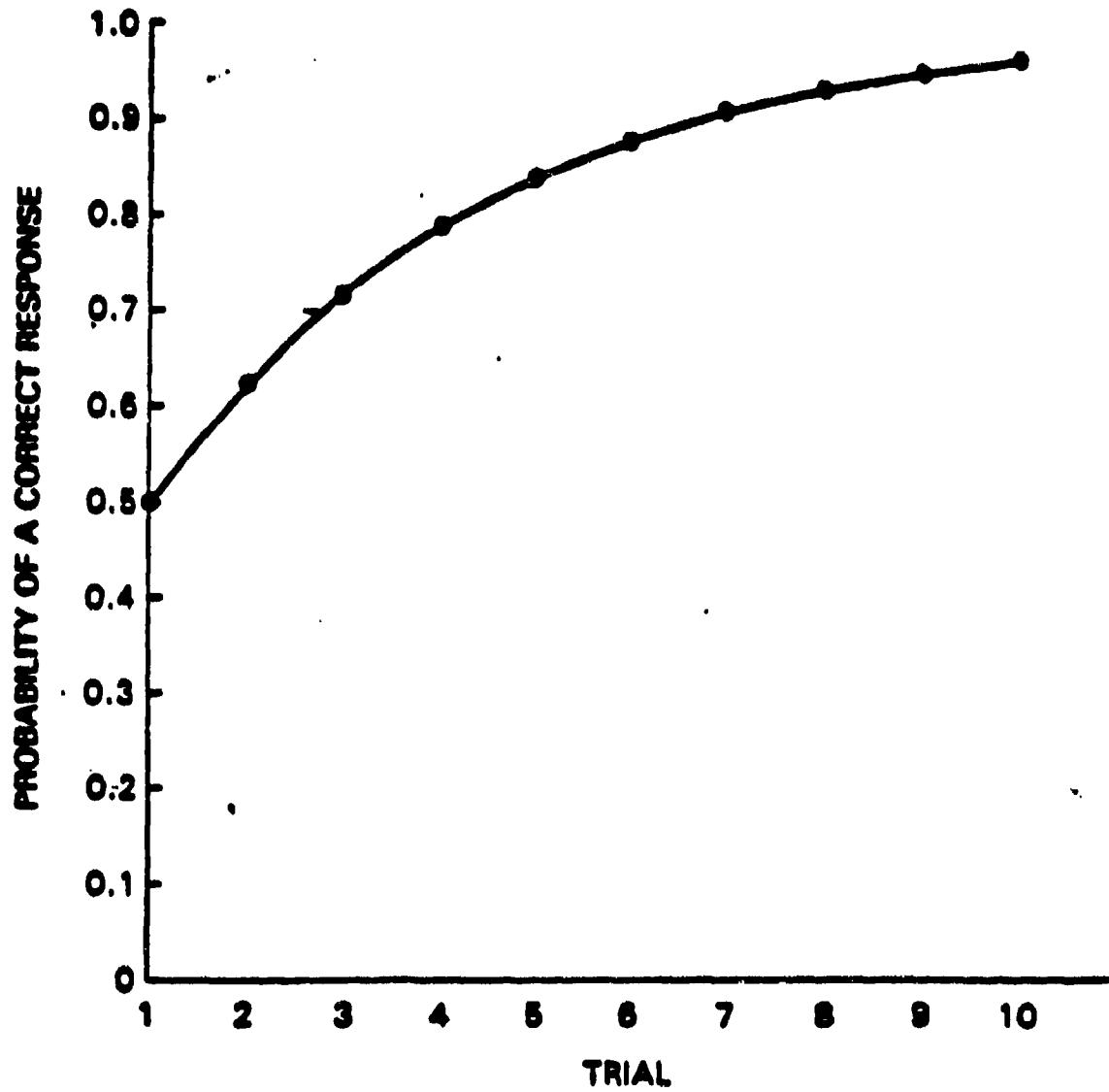


Figure 2. Graph of the probability of a correct response by trial for the linear operator model ($P_1 = 0.5$, $r = 0.25$).

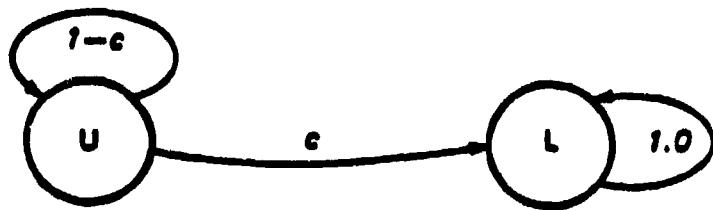
Note that this equation is equivalent to the learning curve for the linear model, equation 6.

The all-or-none and linear models predict the same learning curve but differ greatly in their predictions of the distribution of errors on trials before the last error. In the all-or-none model, no errors occur in the learned state. An error on a trial indicates that the association is in the unlearned state for that trial and for all previous trials. The unlearned state has a constant probability of error ($1 - g$). The all-or-none model exhibits stationarity of the error probability, on trials before the last error, while the linear model does not. The existence of an error gives information regarding the learning state of the student, and this information can be used to optimize the sequencing of instruction.

The fit of the all-or-none model was superior to that of the linear model for simple, paired-associate tasks in a number of empirical studies. For example, Bower (1961) compared the predictions of these two models on a memorization task where syllables were associated with one of two possible responses. The fit of the all-or-none model was impressive and superior to that of the linear model for a large number of sample statistics.

A second strategy for experimentally comparing all-or-none and linear models investigated the basic assumptions regarding behavior on trials before the last error. In the all-or-none model, an error indicates that the association has not been learned. Thus, if a new association is introduced to replace one on which an error has been made, there should be no decrease in performance. This prediction was confirmed by both Rock (1957) and Estes (1960). These results have been subjected to considerable criticism in a controversy summarized by Kintsch (1970). The end result confirms the initial findings regarding the superiority of the all-or-none model.

More complicated tasks, however, show significant deviations between empirical results and the predictions of the all-or-none model. For example, Atkinson and Crothers (1964) found that performance improved on trials just before the last error, when there were more than two response alternatives. In addition, Binford and Gettys (1965) showed that the second guess of subjects was greater than chance and improved with practice--another contradiction of the all-or-none model. A variety of generalizations of the all-or-none model describe learning of more complex tasks. One such model that has experienced some success is the general two-stage learning model (Bower and Theios, 1964). This model has been applied to a number of tasks in which a correct response is not possible on the first trial.



a) Network representation

$$\begin{array}{ccc}
 L_{n+1} & U_{n+1} & P(\text{Correct}) \\
 \begin{bmatrix} 1.0 & 0 \\ c & 1-c \end{bmatrix} & \begin{bmatrix} 1.0 \\ g \end{bmatrix} & \begin{bmatrix} L \\ U \end{bmatrix}
 \end{array}$$

b) Transition matrix and response probabilities

$$\begin{bmatrix} L_0 & U_0 \\ 0 & 1.0 \end{bmatrix}$$

c) Start vector

Figure 3. Network and matrix representations for the one-element model.

The two-stage learning model hypothesizes that associations are in one of three states: an unlearned state in which the probability of a correct response is 0; a partially learned state in which the probability of a correct response is p , with $0 < p < 1$; and a learned state in which the probability of a response is 1. Brainerd, Howe, and Desrochers (1982) reviewed their applications of the model and some of the mathematical techniques used for parameter estimation and model testing.

Rigg and Gray (1980) have applied the one-element model of learning to acquisition and retention data from a complex military procedure. Although the assumptions of the model were clearly violated by the data, the overall fit of the model was good, and the results have some utility for training management.

The experiments described above do not rule out a threshold theory of learning. Underwood and Keppel (1962) showed that a threshold interpretation was consistent with the results of Estes (1960). A threshold model may be found that is equivalent to any all-or-none model (Restle, 1965). However, results such as those of Rock (1957) place constraints on the form the threshold model must have (Restle and Greeno, 1970). The decision between these two modeling approaches to learning must be based on their ability to model other aspects of performance, such as retention and response time, rather than their ability to model acquisition alone.

Instructional Sequencing

Some of the learning models have attempted to predict training differences that arise from differences in instructional sequencing. In general, models in this complex area are not as well developed as the simpler models described above.

Learning Hierarchies. Gagne (1968; 1973; Gagne and Briggs, 1979) developed a practical method for generating sequences of tasks for instruction. The goal was to organize material for students to acquire a progression of higher-order skills. The method was based on the concept of positive transfer from simple to more complex tasks. The simple ones are not just easier, but are also components of the more complex ones. Learning the complex skill, therefore, consists of accumulating the component capabilities through increasing levels of difficulty. Positive transfer results from including the simpler components in the complex tasks.

Learning hierarchies, displayed as diagrams, have boxes showing the successively identified subordinate skills in the task. The hierarchies are generated by asking what capabilities the student must have to perform the task. Gagne (1973) distinguishes between elements that need to be in the immediate learning situation and those that must be retrieved from recall. The latter must be established by previous learning. The instructional sequence uses a succession of learning events to create those capabilities that provide the stimuli from recall essential for learning. The capabilities can not all be established at once. Recall of intellectual skills might require prior learning of subordinate rules and concepts that Gagne calls cumulative learning. "An instructional sequence will be the most effective to the degree that each successive learning event involves a total set of relevant stimuli" (Gagne, 1973, p. 9).

A learning hierarchy, then, is a systematic variation in the stimulus components of a succession of learning events; it has the following characteristics:

1. The hierarchy describes successively achievable intellectual skills hypothesized to contribute substantially to the learning of the target skill and to exhibit positive transfer to it.
2. Learning hierarchies provide descriptions only of intellectual skills, not of verbal information, strategies, motivation, or performance sets. These contribute to learning but are not described specifically.
3. Each node in the hierarchy describes only those prerequisite skills that must be recalled at the moment of learning.
4. A hierarchy is not intended to describe the entire instructional sequence. The stated prerequisite skills must be available to the learner.

The validity of learning hierarchies has been assessed in the areas of mathematics, problem solving, and classification skills (Gagne, 1973). A method is needed, however, for measuring the dependence of skill learning on subordinate skills. Resnick and Wang (1969) applied Guttman's scalogram technique for this purpose and did not find it satisfactory. Validation techniques and measures represent a future research need.

Development of Complex Perceptual-Motor Skills. Research on continuous perceptual-motor tasks, such as those predominant in flying, indicates stages of skill development that might guide sequencing of instruction. Fitts and his colleagues proposed a three-stage model of complex skill acquisition (Fitts, 1964; Fitts, Bahrick, Noble, and Briggs, 1961; Fitts and Posner, 1967). The stages include: (a) A cognitive stage, in which the skill is encoded in sufficient detail to produce a crude approximation of correct performance; (b) an associative stage in which errors are eliminated; and (c) an autonomous stage, in which performance gradually improves further. The first stage relies heavily on cognitive processes. Tests of intellectual ability are good predictors of learning in this stage and research at the University of Illinois Aviation Psychology Laboratory confirmed that the "intellectualization" of flying skills considerably shortened the amount of time needed for solo flight. The learner attends to cues, events, and responses that later go unnoticed. Compared to later stages, demonstrations and verbal analyses are more effective in this cognitive stage. This stage may last for hours or days.

Fitts views complex skill learning as the acquisition of skill in semi-independent subroutines that are performed

successively or concurrently. This view appears well suited to characterize continuous perceptual-motor tasks, especially compared to the forms of task analysis that attempt to divide the skill into discrete stimulus-response events. Fitts further views the overall control of the subroutines by an executive subroutine that initiates and sequences the subroutines for specific skills. Intellectualization in the cognitive stage is the first step in development of the executive program. It allows selection of initial subroutines from preexisting ones and starts creating new ones.

The second stage is variously called intermediate, associative, or fixation. The learner tries existing skills and develops new behavior patterns. The correct patterns are fixated by continued practice, and the probability of errors is reduced to near zero. This stage lasts from hours to months, depending on the complexity of the skill. In flight training, Fitts defines this phase as extending from before the initial solo through granting of a private license, perhaps including as much as the first hundred hours of flying. Critical training issues include schedules of practice (e.g., massed or distributed practice) and training in subroutines (e.g., part-task training in component skills, where invariant subroutines can be identified).

The final autonomous learning stage is characterized by increasing accuracy and speed. Performance is less dependent on external feedback and more dependent on proprioceptive feedback. Cortical associative areas are less involved as learning continues, and control shifts to reliance on lower brain centers (from visual to proprioceptive, for example). After the skills are automated, the learner can perform multiple, competing tasks concurrently (e.g., continue the perceptual-motor task and perform arithmetic problems simultaneously).

Anderson (1982) reformulated Fitts's theory using production systems to describe procedural knowledge. Anderson called the first stage "declarative" and characterized it as encoding of sets of facts to be used later to generate behavior. The learner uses verbal mediation frequently to keep the facts rehearsed. Anderson's theory merges the last two stages into a single procedural learning stage. He acknowledges the gradual conversion from declarative to procedural form by "knowledge compilation," which corresponds to the intermediate stage that Fitts proposed. Although Anderson does not cite it as a separate stage, the two theories are congruent. The supporting research cited by Fitts (e.g., Fitts and Posner, 1967) and by Anderson (1982) addresses both theories.

Research on the development of component skills in continuous perceptual-motor tasks provides additional insights into the sequencing of training. These skills are developed in

identifiable stages whose existence has been replicated (Jaeger, Argwal, and Gottlieb, 1980; Nobel, Trumbo, Ulrich, and Cross, 1966; Trumbo, Noble, Cross, and Ulrich, 1965). Practice at continuous perceptual-motor skills first produces skill in using directional relationships, then skill in timing, and finally skill in using spatial relationships. That is, while beginners respond mainly to displacement error in tracking tasks, experienced operators respond to velocity and acceleration (Briggs, 1961). Temporal organization and coordination are developed through long practice at the task (Lewis, McAllister, and Bechtold, 1953; Pew, 1966).

The existence of these skill development stages suggests that sequencing of training follow a congruent path. One approach is to develop training sequences that progress through training directional relationships to timing to spatial relationships and coordination; the training content should reflect the specific objectives of the task to be trained. Another approach is to speed training by anticipating the next stage; for example, to introduce timing and spatial practice as soon as the student has the directional relationships.

Summary

The model of skill acquisition that would be ideal for application in training system optimization would satisfy three criteria: (a) It would be simple, so that it would not place a computational or assessment burden on the overall analysis system. (b) It would be robust, applicable to a variety of tasks, performance measures, and measures of the amount of practice. (c) It would be consistent with psychological theory; so that model parameters could be estimated from a logical analysis of the information processes required for a task. No modeling framework meets all three of these criteria. However, the power law of practice meets the first two criteria, and current research is investigating the implications of power-function learning on the processes used for skill acquisition. Other models, such as the one-element model, are seriously limited in the scope of their application, or, as is the case for the Hullian model, involve computational complexity both in parameter estimation and in application of the resulting model. Consequently, of the models available, the power law of practice offers the greatest value for a model to optimize training system designs.

Models of Retention

Retention addresses the dynamics of stored information between the time of original learning and the time it is used. Approaches to memory dynamics fall into two of the categories used to describe acquisition models; there are strength and state models of retention, just as there are for learning.

Strength Models of Retention

Strength models of retention functionally describe the strength of an association during a period without practice. Three functions have been proposed to describe this strength: an exponential function, a power function, and an exponential power function decay.

A simple retention function assumes that forgetting occurs at a constant rate; that is,

$$\frac{ds}{dt} = -ks, \quad (8)$$

where s is the strength of the trace, and k is the decay rate.

This representation of forgetting leads to an exponential decay function:

$$s(t) = s_0 e^{-kt}, \quad (9)$$

where s_0 is the strength when $t = 0$.

A number of researchers propose an exponential-decay function to represent decay from short-term memory, where it provides a good account for empirical data (Norman, 1969; Wickelgren and Norman, 1966).

Long-term-memory experiments show systematic deviations from exponential decay (Wickelgren, 1972). A long history of research indicates that memory-decay rate decreases with the age of the memory. One way to accommodate these results is to postulate that decay rate is the product of a constant, k , and a decreasing function, f . If the function f is trace fragility, the equation describing memory decay becomes:

$$\frac{ds}{dt} = -kfs. \quad (10)$$

Equation 10 describes one component of a theory proposed by Wickelgren (1974a). Wickelgren's formulation describes both short- and long-term decay with a single equation. Equation 10 describes the long-term component of the theory. In addition to specifying decay of strength, we must specify a function describing the decay of fragility over time. Following Wickelgren (1974a) we assume that:

$$\frac{df}{dt} = -rf^2, \quad (11)$$

where r is the fragility decay rate.

The assumptions embodied in equations 10 and 11 lead to a power function describing memory decay:

$$s(t) = s_0(1 + rf_0t)^{-k/r}. \quad (12)$$

Wickelgren (1972, 1974b) proposed a similar model of memory dynamics formulated in terms of increasing trace resistance rather than decreasing trace fragility, and the resulting description of the time course of memory is an exponential power function rather than a power function. The decay rate for the strength-resistance formulation is given by:

$$\frac{ds}{dt} = -\frac{ks}{rt^a}, \quad (13)$$

where k is the strength decay rate; and r and a describe the increase of resistance to forgetting.

The resulting decay function is:

$$s(t) = s_0 e^{-\frac{k}{r(1-a)} t^{(1-a)}} \quad (14)$$

Equations 12 and 14 are similar representations of retention processes, differing only in the way trace resistance increases with time (or conversely the way trace fragility decreases over time). In the exponential-power version from equation 13, trace resistance increases as a power function of time. In the power-function version derived from equations 10 and 11, the reciprocal of fragility (analogous to resistance) increases as a linear function of time. Thus, the two models differ only in their predictions about the form of the function describing the increase of resistance to forgetting older information.

The high degree of similarity between these two models makes it difficult to distinguish between them experimentally. Wickelgren (1972) was able to reject linear and exponential-decay models of long-term-memory dynamics. The power and exponential-power models gave comparable fit to the data. However, the parameter estimates for the exponential-power model were easier to interpret than those for the power model, providing some support for the exponential-power model.

Wickelgren (1974a) proposed a hybrid model combining aspects of exponential short-term decay and power long-term decay. A key parameter of his model represents the amount of interference from intervening activities in the retention interval. When interference is high, as is the case when items are represented by a perceptual code, the exponential-decay function dominates. When interference is low, as when information is represented in large semantic chunks, the power function dominates. This model

provides an explanation of memory phenomena without relying on a distinction between short- and long-term memory.

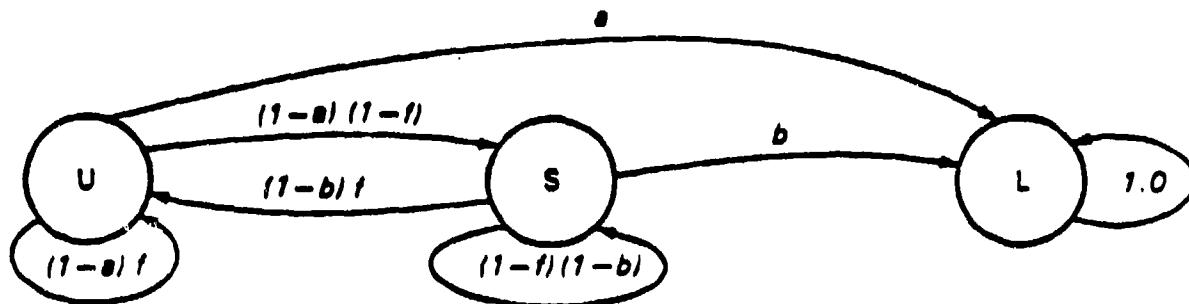
Markov State Models

The all-or-none model can be generalized to account for forgetting, as shown in the three-state model illustrated in Figure 4. The model's three states represent an unlearned state U, a state S in which a transitory memory of the association exists, and a state L in which a permanent representation exists. If knowledge concerning an item is in state S, then a correct response will be given when queried. In addition, the knowledge will move to L with probability b . If the movement to L does not occur, then the item can be forgotten (move to U) with probability f and stay in S with probability $(1-f)$. One interpretation of this model is that state S represents a short-term memory for the association, while state L represents a long-term memory. A closely related model, which is interpreted differently (Restle, 1964; Greeno, 1967), postulates that the states represent ways an individual may code an association that are sufficient to distinguish it from other associations. That is, if codes are sufficient, the association is in state L; if the code is not sufficient, the association is in state S. The all-or-none forgetting model provides a good account of a variety of data. The results are often more easily interpreted by the coding interpretation of the model (see Kintsch, 1970, for a discussion of this issue).

Bower (1967) generalized the all-or-none model to provide a more accurate account for forgetting in the multi-component model. This model represents a stimulus internally by a vector of binary components. The forgetting function may have several alternative forms. In the simplest of these forms, components are forgotten at a constant rate in an all-or-none fashion. These assumptions result in an exponential decay of component information to an asymptotic proportion. Upon further presentations of the association, additional copies of the component values are stored.

Retention of Military Tasks. There has been a significant amount of research investigating the factors affecting the retention of military tasks. Hagman and Rose (1983) have reviewed recent research sponsored by the Army Research Institute, and have stated the following general conclusions on the effect of training, task, and ability variables on skill retention.

1. The level of skill acquisition is a major determiner of retention. Increasing the number of repetitions of a task during training will increase later retention (Block and Burns, 1976; Goldberg, Drillings, and Dressel, 1982; Schendel and Hagman, 1980). Repetition is generally effective when it applies to both initial practice trials and test trials.



a) Network representation

	L_{n+1}	S_{n+1}	U_{n+1}	$P(\text{Correct})$
L_n	1.0	0	0	[1.0]
S_n	b	$(1-b)(1-f)$	$(1-b)f$	[1.0]
U_n	0	$(1-a)(1-f)$	$(1-a)f$	[a]

b) Transition matrix and response probabilities

Figure 4. Network and matrix representations of an all-or-none learning model with short-term forgetting.

However, repeated testing does not enhance retention when the task is performed with a job aid.

2. Retention is enhanced by active practice (Hagman, 1980a; Holmgren, Hilligoss, Swezey, and Eakins, 1979) and spaced practice (Hagman, 1980b).
3. Use of mnemonic techniques does not necessarily enhance retention.
4. Procedural tasks are forgotten much more quickly than continuous control tasks (Schendel, Shields, and Katz, 1978). Among procedural tasks, forgetting is best predicted by the number of steps in the task. Steps that lack cues from the equipment, are unclear to the soldier, are passive, and first and last steps are forgotten relatively quickly (Osborn, Campbell, and Harris, 1979).

5. General ability affects task acquisition rather than retention. That is, high-ability trainees will learn a task faster than low-ability trainees, but if both groups are trained to the same performance standard, they will exhibit equal retention.

Rose, Czarnolewski, Gragg, Austin, Ford, Doyle, and Hagman (1985) have developed a predictive model that summarizes many of the empirical results relating to skill retention. The model postulates an exponential retention function based on a retention index that combines ratings of the following ten attributes.

1. The existence of job or memory aids to be used in performing the task,
2. The quality of the job or memory aid,
3. The number of steps required to perform the task,
4. The extent to which the steps must be performed in a definite sequence,
5. Whether the task provides feedback on whether it is being performed correctly,
6. Whether there is a time limit,
7. The cognitive requirements of the task,
8. The number of facts, terms, names, rules, or ideas that the soldier must memorize to perform the task,
9. How hard the facts, terms, names, etc. are to remember, and
10. The motor skill demands of the task?

The retention index is an additive combination of scale values that depend on the responses to the ten questions. The scale values were determined using multiple regression.

The model provides a reasonably accurate prediction of retention over a wide retention interval (Rose, Czarnolewski, Gragg, Austin, Ford, Doyle, and Hagman, 1985). In addition, the data requirements are moderate, requiring between five and eight minutes effort by subject-matter experts per task (Rose, Radtke, Shettal, and Hagman, 1985). Thus, the model provides a good account of retention of a variety of military tasks.

Summary

Although skill retention is a critical concern for unit training, it is not as important in institutional training.

Nevertheless, the retention literature contains information that can be used to provide training in a training institution that maximizes later retention in an operational unit. Some of the relevant findings in the literature are the following.

1. Perhaps the most important finding regarding retention is that initial learning is a major determinant of later retention. Thus, it is important that the most critical tasks be trained initially to a high level to guard against later forgetting.
2. The research has uncovered task differences in retention, with procedures being forgotten much more quickly than other tasks.
3. The modeling literature indicates that older memories last longer than newer memories of the same strength. In that case we may want to make sure that the most critical tasks are taught early in the training history.

When we become concerned with broader training issues that encompass both institutional and unit training, then issues of retention will become much more important. Many of the modeling constructs required to address retention are currently available.

Models of Transfer of Training

For training to be effective, the trainee must be able to apply the knowledge and skills obtained in the training setting to the operational setting. Thus, transfer of training forms the basis of the overall assessment of the effectiveness of a training system.

Transfer of training is an inherently complex issue. The transfer of training from a training device to actual equipment is dependent upon the fidelity with which the training device represents the operational environment, the similarity of the skills taught on the training device to those required in the operational environment, and other factors. Because of the complexity of transfer of training, we are faced with a situation in which there are multiple measures of transfer of training, few theoretical treatments, and limited predictive capability.

Measures of Transfer

In the simplest of transfer of training designs, transfer of training is measured by comparing the performance on a transfer task between two groups. The experimental group receives prior training on a training task; the control group does not receive this training. Several different measures of transfer have been proposed. We distinguish two classes of transfer measures: (a) measures based on comparisons of the performance on the transfer

task between the experimental and control groups, and (b) methods based on a measure of savings of training on the transfer task produced by the training on the training task.

The simplest of the measures of the first kind describes transfer as follows:

$$\frac{E - C}{C} \times 100 \quad (15)$$

where E = the performance of the experimental group on the transfer task, and

C = the performance of the control group on the transfer task.

Other measures use the same numerator as in equation 15, but have the denominator of either $T - C$, where T is the maximum possible performance on the transfer task, or $E + C$. The measure of transfer with the denominator $E + C$ has the advantage that the resulting measure is always between -100 (for maximum negative transfer) and +100 (for maximum positive transfer).

A seminal paper by Roscoe (1971) introduced measures of transfer effectiveness based on savings, such as the cumulative transfer effectiveness ratio (CTER):

$$\text{CTER} = \frac{Y_0 - Y_x}{X} = \frac{\text{Savings in Aircraft Training Time}}{\text{Simulator Training Time}} \quad (16)$$

where Y_0 = time, trials, or errors required by a control group to reach a performance criterion;

Y_x = corresponding measure for simulator-trained group;

X = time, trials, or errors by the simulator-trained group during simulator training.

The CTER is actually a decreasing function of X , tending toward zero for large X . By estimating the CTER for various values of X , and by considering associated costs, early studies in simulator training economics were conducted (e.g., Holman, 1979; Provenmire and Roscoe, 1973).

Theories of Transfer

A theory of transfer must, at the least, relate the degree of transfer to the characteristics of the two settings. In particular, a theory of transfer of training must predict operational setting performance as a function of the performance criterion in the training setting, and of the differences between the two settings.

Early theories of transfer of training (Thorndike, 1903) postulated that if two tasks had the same aims, elements, or approaches, then training on one task would transfer to the other task. This theory, termed the theory of "identical elements," is significant in that it postulates that transfer is related to similarity of specific elements of the task, whereas earlier theories of transfer proposed general transfer mechanisms. However, the identical-elements theory is not expressed as a mathematical model of transfer. Furthermore, the identical-elements theory does not provide an account for negative transfer, which reliably occurs under certain conditions.

The most well-known approach to transfer of training is the stimulus-response analysis of Osgood (1949). Osgood developed a relationship that relates transfer of training to stimulus and response similarity of tasks. Transfer is maximal when both the stimuli and responses are similar. When response similarity is high, increasing stimulus similarity leads to increasing positive transfer of training. However, when response similarity is low between the two tasks, or responses are antagonistic, increasing stimulus similarity leads to increasingly negative transfer of training. The greatest negative transfer occurs when antagonistic responses are associated to the same stimuli. Conceptually, this transfer surface can be represented by the multiplicative function,

$$T_{12} = S_{12} (R_{12} - k), \quad (17)$$

where T_{12} = the transfer of training from task 1 to task 2;
 S_{12} = the stimulus similarity;
 R_{12} = the response similarity; and
 k = the level of response similarity that produces neutral transfer of training.

Current psychological theory does not address transfer of training directly, but some theories of knowledge acquisition make predictions regarding mechanisms by which transfer may occur. These theories attempt to identify the mental model by which an individual represents skills and knowledge that are learned. Recent work in this area, such as the work of Kieras (1985), may have implications on the prediction of transfer of training, but it will be some time before such cognitive theories are sufficiently advanced to be used as the basis of methods for training-system optimization.

Use of Transfer to Allocate Tasks to Training Devices

As aircraft simulators became more effective -- and expensive -- training managers are faced the question of how much time a student should spend on the simulator before moving on to actual in-flight instruction. The studies reviewed here address various aspects of this issue of efficient simulator use. None

of these papers is involved in simulator design, though many of the results will prove useful in addressing that area also.

The research of Bickley (1980), Carter and Trollip (1980), and Cronholm (1985) discussed below, was based on the foundations laid by Roscoe (1971), though each of these researchers elected to focus on different measures of effectiveness.

Minimizing training cost. Bickley's (1980) research involved the use of empirically fit iso-performance curves for optimizing simulator usage.

Bickley postulated an exponentially decaying iso-performance curves of the form:

$$y = a e^{-bx} + c \quad (18)$$

where y = training in aircraft required to reach performance criterion after simulator training

x = simulator training

a , b , c = positive constants, parameters of the model.

Bickley verified that this formulation is consistent with previous empirical data (Provenmire and Roscoe, 1973). He then conducted an ambitious program of empirical research on training with a prototype AH-1 Cobra helicopter flight simulator (AH1FS) and the aircraft itself. Thirty-one tasks, both procedural and psychomotor, were investigated. The decaying exponential form was found consistent with the data collected for each of these tasks. However, for several of these tasks simulator training was more effective than anticipated in the experimental plan, so that all the data fell in the asymptotic region of performance; in these cases there was a dearth of data for fitting the iso-performance curves in the region of greatest interest. Bickley acknowledged that other forms might also prove consistent with the data.

Bickley then went on to incorporate these empirical iso-performance curves into a simple model for total training cost. This total cost model just adds the costs attributed to simulator use to the costs attributed to aircraft use. These costs are each calculated as the respective student time spent on each medium multiplied by cost per time for that medium; i.e., the simplest linear cost model for each medium. (Bickley does not address the source of the cost rate data.)

Using elementary calculus, he developed an expression for determining the optimal simulator training time for each task. "Optimal" here refers to minimizing costs for training to criterion. Given his assumptions, cost is minimized when the amount of simulator training x satisfies

$$x = (\ln C_A + \ln a + \ln B - \ln C_s)/b \quad (19)$$

where C_A = the cost rate for aircraft training,
 C_s = the cost rate for simulator training, and
a and b are previously identified constants defining
the simulator/aircraft iso-performance tradeoff curve.

In summary, Bickley's research serves as a useful introduction to the simulator/aircraft tradeoff problem and provides a useful empirical starting point for further research.

His theoretical results are limited, however, to cases in which the iso-performance curves have a very specific functional form (negative exponential) and costs are linear with simulator and aircraft usage.

Maximization of training effectiveness. Whereas Bickley concluded his paper with a simple cost minimization methodology, Carter and Trollip (1980) present a simple performance maximization methodology. They address the question: when should simulator-to-aircraft transfer occur in order to maximize terminal performance, given a fixed training budget?

For purpose of exposition the authors employ simple hyperbolic iso-performance curves, of the form:

$$xy = c \quad (20)$$

where y = training on aircraft to bring student to criterion performance, after simulator training completed;
x = training on simulator;
c = constant, dependent on simulator, aircraft, and final student performance in the aircraft.

The constant in this equation has no simple physical or operational interpretation. However, any particular level of final student performance will have an associated value of c; the higher the level of performance, the higher c. Any combination of x and y satisfying $xy = c$ will yield the same final student performance as any other values of x and y satisfying this equation for the same value of c.

The authors proceed to illustrate the performance maximization problem graphically and then to solve it using the Lagrange multiplier technique -- a standard approach to constrained optimization problems. The approach is used to develop a general solution to the problem:

Maximize $f(x,y) = xy$ (i.e. the associated terminal student performance level) subject to a particular budget constraint:

$$a_x x + a_y y = b,$$

where a_x and a_y are costs per hour of simulator and aircraft training, respectively, and where b is the total per-student budget in dollars. If every simulator hour cost \$10/hour and every aircraft hour cost \$20/hour, and a total budget of \$140 per student is available, then this constraint equation becomes:

$$10x + 20y = 140.$$

Any number of simulator-aircraft training hour allocations would satisfy this budget (e.g., 14-0, 10-2, 6-4, 2-6, and 0-7). The Lagrange multiplier technique determines which of these combinations (or any other along the iso-budget constraint) results in the greatest final student performance y . In this case the maximum performance is achieved when $x = 7$ and $y = 3.5$, corresponding to the iso-performance curve $f(x,y) = xy = 24.5$.

In summary, this paper is a straightforward application of the classical Lagrange multiplier optimization technique to a training trade-off problem. It is quite limited in scope, developing a particular example involving iso-performance and iso-budget curves of the simplest sort of mathematical form. It does not address the general applicability of this formulation to more general iso-performance and iso-budget curves. Further, the problem addressed by this paper, maximizing performance gains within a fixed training budget, may prove to be of less interest to the training community than the complementary problem of minimizing training costs incurred to achieve a criterion level of performance.

An approach to optimize cost and effectiveness. In this paper Cronholm (1985) generalizes the training cost minimization problem and the training performance maximization problem.

He provides a deliberate and general mathematical development of the skill-defined task sequence optimization problem. For the two-task problem (which he later generalizes to n tasks) he breaks the training process into three steps: (a) Task 1 (Simulator) Training, (b) Task 2 (Aircraft) Training, and (c) Transfer of Skill from Task 1 to Task 2. He represents each of these processes in terms of mathematically general learning and learning transfer functions. In particular he assumes only that these learning curves and transfer functions are monotone increasing, continuous, and differentiable. The strict monotonicity assures that the learning curves have inverses, which is important to the development of Cronholm's theory. Cronholm's assumptions are met intuitively in most situations, or can be restricted to these assumptions without impacting any real options.

A third kind of mathematical function is also required in Cronholm's formulation: cost curves. He follows Bickley in assuming that the cost incurred at each learning stage is a simple linear function (time spent x cost per hour) with no setup (investment) step in making transition from zero to positive resource allocation at a particular task. The mathematics of his approach make explicit use of this assumption.

Cronholm employs the calculus to optimize training, first for the cost-minimization problem and second for the performance maximization problem. In a bit of elegant mathematics he finds that the key issue -- when to program student transfer from simulator to aircraft -- is resolved identically for each of these problems. The only difference between the optimal solutions under cost-minimization and performance maximization is that in the former case aircraft training is halted when criterion performance is achieved, while in the latter case aircraft training is continued until budgetary resources are exhausted.

Cronholm's results may be summarized as follows. The optimal training resource allocation to task 1 will be given by the following sequence of equations (where the " $^{-1}$ " refers to the inverse of the associated functions):

$$x_{1\text{opt}} = u^{-1}(r) \quad (21)$$

where

$$r = \frac{c_1}{c_2} = \frac{\text{Cost rate on task 1}}{\text{Cost rate on task 2}}$$

and

$$u(x_1) = \frac{d}{dx_1} g^{-1}(t[f(x_1)])$$

is an intermediate function based on the functions f , t , and g :

$y_1 = f(x_1)$ = learning on task 1 given resource allocation x_1 to task 1

$y_{20} = t(y_1)$ = transfer of learning to start of task 2, and

$y_2 = g(x_2 + g^{-1}(t[f(x_1)]))$ = learning on task 2 given resource allocation x_2 to task 2.

(Here $g(\cdot)$ represents a learning curve and $g^{-1}(t[f(x_1)])$ represents a head start on task 2 resulting from investment x_1 to task 1).

Cronholm goes on to illustrate the behavior of the solution when certain specific mathematical forms are substituted, viz.

negative exponential learning curves and a linear transfer function. His examination includes a parametric analysis.

Two potential complications in applying Cronholm's method deserve to be mentioned. First, if the media cost functions are linear but involve a setup cost then the setup cost must be added in before or after the Cronholm optimization. If it is desired to determine whether a medium should be used, the problem must be solved first with the medium in place (thereby obtaining total training plus setup costs) and then without the medium in place (obtaining total training and setup costs for active media only).

Second, Cronholm's assumptions concerning the nature of the learning and transfer curves do not appear sufficient to guarantee that the $u(x_i)$ function is monotone and therefore invertible. The condition

$$u(x_i) = r$$

is only necessary, not sufficient, for optimality. It is possible that more than one value of x_i will be found to satisfy this condition.

In summary, Cronholm's paper provides a synthesis and extension of the work of other researchers. When his assumptions concerning learning, transfer, and cost curves hold true, his method provides the foundation for an extremely efficient algorithm for solving the single-skill-training resource allocation problem.

Prediction of Transfer of Training for Military Tasks

The Training Device Effectiveness Model (TRAINVICE) was developed by Wheaton, Rose, Fingerman, Korotkin, and Holding (1976) to predict transfer of training from a training device to operational equipment. Several versions of the TRAINVICE model have been developed since that time (Narva, 1979; Swazey and Evans, 1980). These versions differ in the details of their analysis, but all share most of the characteristics of the original formulation, which we will describe here. (For a discussion of the differences between versions of TRAINVICE, see Tufano and Evans, 1984; Knerr, Nadler, and Dowell, 1984.)

The estimates of the TRAINVICE model are based on a set of ratings for each subtask performed on the training device. The overall estimate of transfer of training is based on the sum of an estimate for each subtask. The TRAINVICE model estimates transfer of training for each subtask as a product of the following four factors.

1. Task communality. This factor measures the degree of overlap between the tasks performed on the training device and the tasks performed on actual equipment.
2. Similarity. This factor rates the physical and functional similarity between the training device and the actual equipment. Physical similarity is a measure of how well the displays and controls are represented on the training device. Functional similarity is a measure of how similar the information processing activities required to perceive and operate the displays and controls on the training device are to the corresponding activities on actual equipment.
3. Training Techniques. This factor rates how well the training device implements the appropriate learning guidelines, considering the type of skill required in each subtask.
4. Learning Deficit. This factor rates the difference between the trainee's entry skill level and the level of skill that is required to perform the subtask.

The TRAINVICE model has been used to evaluate several training devices (e.g., Wheaton, Rose, Fingerman, Leonard, and Boycan, 1976; Harris, Ford, Tufano, and Wiggs, 1983; Klein, Kane, Chinn, and Jukes, 1978). A typical finding in all applications is that the model does not distinguish between different training-device designs. Most applications produce an estimate of moderate transfer independent of the training-device design. Rigorous validations of the TRAINVICE model have not been conducted.

The relative insensitivity of the TRAINVICE model to training-device design variables reflects both the characteristics of the model and the difficulty of estimating a variable that is as complex as transfer of training. Since the model is additive over subtasks, and since a training device is likely to provide effective training on some subtasks, and ineffective training on others, we might expect moderate estimates of transfer to be common. However, the results of this model indicate that we have only limited ability to predict transfer of training.

Cost Estimation

To support the development of cost-effective training systems, it is essential both to understand and evaluate the costs of current training systems, and to forecast the costs of future training systems. Our review indicates a wealth of literature addressing the first activity and a dearth addressing the second. More accurately, there are many reports on training system cost modeling: cost categorization, cost aggregation, cost proration, and life cycle costing. Yet the only literature which we have found on training system cost forecasting is that involved with long-range forecasting methodology in general. Similarly, data on costs of current training systems and training devices are available in some reports, but long-range forecasts do not seem to see publication. The literature identified in Table 5 is representative of what is available.

Cost Modeling: Current Training Systems

Training system and training device cost analyses in the past have used many different, often ad hoc, cost classification schemes. This has made it difficult to use these analyses and the associated data to compare different training systems, devices, and programs. To rectify this problem Office of the Secretary of Defense for Research and Engineering (OSDR&E) contracted with the Institute for Defense Analyses (IDA) to develop a standardized cost element structure for defense training (Knapp and Orlansky, 1983). The resulting cost element structure (CES) is indeed comprehensive. It was developed using a work breakdown structure approach. Though the focus of this effort was flight training, it appears to be applicable to other weapon systems and training programs through a simple relabeling of categories. Knapp and Orlansky provide a cost element structure for training, but do not provide associated aggregation, proration, or life cycle costing methods. However, the use of such methods with this CES is apparent in a subsequent IDA study on the operating costs of aircraft and flight simulators (Orlansky, Knapp, and String, 1984). An earlier study by String and Orlansky (1977) also describes the kind of cost modeling built on such a cost categorization scheme in order to conduct analyses, in this case relating to the cost-effectiveness of flight simulators for military training.

A 1980 Cost and Training Effectiveness Analysis Performance Guide report presents a cost model suitable for manual and hand calculator employment (Matlick, Rosen, and Berger, 1980). This "model" is actually more a straightforward tutorial on cost estimation, progressing by example. It offers useful guidance for the cost analyst (e.g., reminding him to spend the greatest effort on those cost elements contributing the largest absolute uncertainty to the total system, and showing him how to develop

Table 5

Relevant Literature for Long-Range Training System Cost Forecasting Methodology

<u>Report</u>	<u>Cost Modeling</u>	<u>Cost Forecasting</u>
Allbee & Semple (1981)	Aircrew TS/TD cost categori- zation, proration, LCC ^{**} , and data	
Armstrong (1985)		Comprehensive textbook on long range fore- casting in general
Armstrong (1986)		Review of forecasting methods
Knapp & Orlansky (1983)	Training cost categorization	
Martino (1983)		Textbook on technolo- gical forecasting in general
Marcus, Patterson, Bennett & Gershman (1980)	TS/TD cost cate- gorization, aggre- gation	
Matlick, Rosen, & Berger (1980)	TS/TD cost cate- gorization, aggre- gation	
Orlansky, Knapp, & String (1984)	Acft/simulator aggregated costs	Cost trend analysis
String & Orlansky (1977)	ACFT TS/TD cost categorization and data needs	
Thode & Walker (1983)	ACFT TS cost cate- gorization, aggre- gation	

^{**}TS/TD - training system/training device
** LCC - life cycle costing

estimates based on task and system similarity). It provides datacollection work sheets. Examples are constructed around artillery training.

A more formal cost model focused on aircrew training devices was prepared for the Air Force Human Resources Laboratory (Allbee and Semple, 1981). It addresses cost proration in considerable detail. It also provides a sophisticated life cycle cost model geared toward use of available Air Force accounting system data to determine the cost of aircrew training and to differentiate between simulator and flight training costs. A somewhat similar but much less detailed model is presented by Thode and Walker (1983). Yet another cost model is embedded in the cost-effectiveness methodology developed by Marcus, et al., (1980), and discussed elsewhere in this review. It is fully automated as part of a larger computer model.

Cost Forecasting: Future Systems

None of the above studies and models addresses the cost of future, yet-to-be-built training devices. Further, little or no attempt is made to cost a training system in terms of constituent parts (motion system, visual system, computer, etc.). This is presumably true in part due to the difficulty of obtaining such data. Knapp and Orlansky (1983) point out that training equipment is often procured via firm fixed-price (FFP) and fixed-price incentive-fee (FPIF) contracts which provide the Services little leverage in the specification of cost detail. For the cost modeling associated with current training devices, such detail is not really needed. But the estimation of costs of future training devices, with capabilities beyond those of current devices, requires system disaggregation and subsystem cost forecasting.

Discussions with a few training device developers and Service procurement specialists suggest that cost projection for sophisticated training devices, such as flight simulators, is largely a matter of expert judgment. Practitioners in forecasting acknowledge the need to rely on expert judgment, but offer guidelines on a more structured approach (e.g., Armstrong, 1985, 1986; Martino, 1983). In particular, Armstrong recommends disaggregation and then the use of a several forecasting methods, which are then combined. Some elements of the training system are better forecast with one method than another. For instance, simple extrapolation is quite valid for such things as the cost of floor space or a simulator mechanical motion subsystem, where technology is either unimportant or unlikely to change significantly. For other subsystems, such as the computer image generation (CIG) component of a simulator, the technology is evolving at a very rapid pace, so expert judgment may be most appropriate. When expert judgment is to be employed, techniques such as the Delphi process have proved particularly valuable.

We have not identified literature devoted to forecasting the costs of future training devices, and it is not within the scope of this review to embark on a description of the larger forecasting literature. However, Armstrong (1986) presents a good survey of forecasting methods, and the textbook by Armstrong (1985) is a comprehensive presentation on long-range forecasting methods. In addition, Martino (1983) focuses on technological forecasting, and provides a particularly good description of the use of the Delphi technique.

Estimation Procedures for OSBATS Cost Data

One of the major concerns of a study by Willis, Guha, and Hunter (1988) to investigate data collection and utilization procedures for the OSBATS model was the development of procedures to estimate the cost of existing training devices, as well as to predict the cost of fidelity dimension levels and instructional features. They developed procedures that combined the use of existing data with cost estimating relationships.

Cost elements for existing training devices were estimated using both available data and cost estimating relationships. For example, contractor system engineering costs for development of training device were obtained by examining the proposals for awarded contracts. SME analysis of these proposals were used to estimate other cost elements, such as front-end analysis costs and research and development costs as a percentage of the contractor system engineering costs.

Costs of instructional features and fidelity dimension levels were estimated using the Constructive Cost Model (COCOMO) by Barry Boehm (1981). The COCOMO model is designed to estimate software development costs as a function of the size of the project. Since a large percentage of the costs for instructional features and fidelity levels represents software development, Willis, et al. (1988) determined that the COCOMO model was appropriate. The estimates of the COCOMO model were spot checked against instructional features and fidelity levels for which the number of lines of source code were known, and the overall cost could be estimated with relatively high accuracy. The checks of the model estimates determined that the values estimated by the model were in the same range as those estimated from the number of lines of source code.

Research Plan

Current research knowledge provides a framework for optimizing the design of training systems. However, the research does not provide us with the specific knowledge required to estimate critical model parameters, such as learning rates and transfer-of-training functions. In place of actual data or validated theory, we have made some general assumptions about the nature of these functions. For example, learning and transfer functions are assumed to be power functions with parameters based on subjective judgments and hypothesized relationships. Some of these assumptions are central to the OSBATS model.

Areas where we lack data relate directly to the input requirements of the model, and to the model processes that transform the input data into recommendations for training device designs. We may specify the research requirements by examining these data and processes to determine what research questions must be answered to improve the quality of input data and the accuracy of model processes. Analysis of model output can indicate the relative importance of different research questions. It is more critical to know the answer to questions that have a large impact on the recommendations of the model, than it is to be able to answer questions for which the model recommendations are relatively insensitive.

Each of the relevant questions, or research topics, can be answered to varying degrees by adjusting the research effort and expense dedicated to it. The research plan we have developed uses a resource-allocation model -- similar to the one developed for training-device design -- to maximize the benefit/cost ratio of answering a set of questions. The results of this model specify the optimal level of effort to dedicate to each research topic as a function of the total budget for the research effort. Thus, the model identifies those specific research projects in which a substantial improvement in the quality of the model may be obtained for a relatively small effort.

This report presents the second version of the research plan. The first version of the plan was produced before the model software had been developed, and is documented by Young, Luster, Stock, Mumaw, and Sticha (1986). This plan refines the original research plan by incorporating knowledge that was gained from the implementation and evaluation of the OSBATS model. Although we used the original research plan as one source for the current plan, we have completely redefined many of the research topics and specific research projects.

The remainder of this plan describes the research topics that were considered, the general analysis procedure, the rationale for assessed costs and benefits, and the model results.

Defining the Research Topics

The research topics addressed in this plan concentrate on basic psychological research to enhance our knowledge of learning and transfer processes and their relation to task and training-device characteristics. Our orientation towards psychological research implies that the plan does not address issues related to estimating the cost of training devices and the training programs of which they are a part. Cost estimation represents one area where additional research is needed. However, we chose to focus the research plan on psychological research.

To ensure the comprehensiveness of the list of research topics, we developed a framework that specified the kinds of factors and interactions that must be understood to maximize the validity of the OSBATS model. The framework specifies three classes of factors.

1. Device factors. These factors describe the characteristics of a training device that make it train more efficiently or effectively. We consider two types of device factors, instructional features and fidelity levels.
2. Student factors. These factors describe the skills and abilities of the students relevant to the training requirements. Two types of factors are considered, student aptitudes and specific relevant experience.
3. Task factors. These factors describe the characteristics of tasks that mediate the device requirements. Specifically, task factors include information-processing characteristics, cue and response requirements, and overall task difficulty.

The critical relationships which must be captured by the OSBATS model associate changes in the device, student, and task factors with changes in the following two critical dependent variables: learning rate, and transfer of training. Some of the most critical of these interactions are the following:

1. Task difficulty and learning rate,
2. Student aptitude and learning rate,
3. Task cue and response requirements, device fidelity and transfer of training, and
4. Task information-processing characteristics, device instructional features and learning rate

The framework was used, along with other guides to develop a list of research topics. We used the following sources, in addition to the framework, to generate the research issues: (a)

the list of data variables, (b) the original research plan, (c) our knowledge of the model assumptions, and (d) the results of sensitivity analyses.

The Research Topics

The following twelve research topics were generated using this process.

1. Performance measurement methods. The goal of this research topic is to develop consistent, criterion-referenced methods to measure performance on a variety of tasks on a common numerical scale.
2. Task evaluation factors. Task evaluation factors are used to evaluate the need for and benefits from training in a simulated environment. The goal of this research topic is to generate a comprehensive set of task evaluation factors and to specify how ratings on these factors should be aggregated.
3. Task cue and response requirements. Task cue and response requirements are currently determined using a rule base that works in a limited domain of tasks and fidelity dimensions. The goals of this research topic is to develop both a comprehensive set of cue and response dimensions and general procedures for selecting the appropriate dimensions for any training domain.
4. Task training hours. The goal of this research topic is to develop methods to estimate the training time required to achieve the performance standard on a new task for which no training data exist.
5. Task characteristics/instructional features. Current procedures address instructional feature requirements in a single training domain. The goal of this research topic is to generalize the relationships to other domains.
6. Fidelity dimensions. The goal of this research topic is to develop methods to determine fidelity requirements over a wide variety of tasks and training domains.
7. Instructional features. Current model procedures assume that instructional features are either present or absent in a training device. The goal of this research topic is to develop a framework that considers different levels at which instructional features may be implemented, and to develop procedures that determine the level of instructional feature required by any task.

8. Learning assumptions. The goal of this research is to evaluate some of the specific assumptions about the learning process made by the OSBATS model.
9. Aptitude mixture. Current model procedures do not consider the distribution of aptitude of the student population in making their recommendations. The goal of this research topic is to develop model procedures that are sensitive to differences in student aptitude.
10. Instructional features/fidelity combination. The goal of this research topic is to develop methods that specify the relative value of instructional features and fidelity features in a training-device design.
11. Model advisor. The goal of this research topic is to develop automated methods to explain results, suggest further analyses, and incorporate confidence values into the model recommendations.
12. Prerequisite skills. Current procedures focus on the specific tasks that need to be trained, and do not consider whether the student possesses the required prerequisite skill necessary to perform these tasks. The goal of this research topic is to determine whether an analysis of prerequisite skills would enhance the capabilities of the OSBATS model.

Possible Levels of Research

We developed several research options to address each of the research issues. The options varied both in cost and in the extent to which they closed the knowledge gap regarding each of the issues. In general, the options are cumulative; that is, the research in the more expensive options builds on the results of the less expensive options. The total set of options considered by the analysis is shown in Figure 5.

In this section, we describe the options for each research topic. The description begins with a summary of the current knowledge and the research need. Then, each level of effort will be described, and the rationale for cost and benefit assessments will be outlined. The numerical estimates for cost and benefit will be described in the following section. The first level of research is shown on the first box, and so forth.

Topic 1: Performance measurement methods. The student entry performance and task performance standard were assessed on a numerical scale based on subject-matter expert judgments. Subsequent discussions with the judges made it clear that the performance values that were assessed were relative values, with the performance standard set somewhat arbitrarily at 70%, and the

VARIABLE	1	2	3	4
1 Perf Measure Me	Perf Measure Su	Normed Ref Msrmnt	Criterion Ref Msrmnt	
2 Task Evaluatn F	Compile Dictionary	Develop Analyt Met		
3 Task Cue/Resp R	Cue/Resp Generation	Develop Taxonomy	Dim Selection	Anal for New Domain
4 Task Training H	New Task Prediction			
5 Task Char/Inst	Establish Framework	Spec Knowldg En		
6 Fidelity Dimensions	Charact Known Relt	Conduct Res Progra		
7 Instructional Features	Develop Taxonomy	Feature Select Met	Charac Known Rela	Conduct Research P
8 Learning Assumption	Test Current As	Task Type Framework	Task Type Research	IF/Fidelity Effects
9 Aptitude Mixture	Charact Relations	Model Concept De		
10 IF/Fidelity Combo	Analytical Evaluatn	Develop Concept De		
11 Model Advisor	Explain Results	Boundary Cases	Suggest Sen Analys	
12 Prerequisite Skills	Analyze Relevance			

Figure 5. Research topics and their levels.

entry performance level proportionately lower. The original intention of the model was that expert performance would receive the score 1.0, performance of untrained soldiers would receive the score 0.0, and intermediate levels of performance would receive a score in proportion to the level of performance. However, implementing this scoring procedure over a wide variety of tasks requires scaling and measurement research. Three levels of research were considered to develop performance measurement methods.

The first level of research (Performance Measure Survey) would be a survey of performance measurement methods currently used in military training applications. The results of this survey would be a catalog of measurement methods currently in use. We could then select or modify existing methods, where appropriate, to use as to obtain student entry and standard performance estimates.

The second level (Normed Reference Measurement) builds upon the first by development of norm-referenced measurement methods. This task requires a substantial effort because of the wide variety of tasks that must be covered by the methods. Norm-referenced methods would support most OSBATS analyses. However, it would not be possible to obtain accurate estimates of the overall cost to meet training requirements.

The third level (Criterion Referenced Measurement) extends the previous levels through development of criterion-referenced measurement methods. These methods would make more accurate training cost estimates possible. In addition, they would provide a straightforward procedures for the OSBATS model user to investigate the effects on changes in selection criteria or performance standards.

Topic 2: Task evaluation factors. The OSBATS model currently contains a list of six factors that are used to represent the non-monetary benefits of simulation. These factors are specific to the aviation problem used as the initial example of the model. In addition, the methods used to combine ratings for different factors will need to be revised as additional factors are included.

The initial level of research (Compile Dictionary) on this topic compiles a comprehensive dictionary of task evaluation factors. This dictionary would list the task domains for which each evaluation factor is relevant.

The second level of suggested research (Develop Analytic Methods) would develop the analytic methods that could aggregate task ratings on the task evaluation factors to produce an overall index of simulation benefit.

Topic 3: Task cue and response requirements. The current version of the OSBATS model considers eleven fidelity dimensions that form the basis of task cue and response requirements. The fidelity dimensions depend on both the task domain and the type of training device being designed. For example, the fidelity dimensions for armor turret maintenance are considerably different from those for rotary-wing operations (Sticha, Blacksten, Bueda, Singer, Gilligan, Mumaw, & Morrison, 1988). In addition, the dimensions that are most relevant for expensive, full mission simulators are likely to be different from those relevant to part mission simulators. The goal of this research would be to generate a comprehensive set of fidelity dimensions, organize these dimensions, and develop procedures that would specify the appropriate dimensions for any specific training-device design problem.

The first level of research (Cue/Response Generation) considered for this topic would generate a set of fidelity dimensions and levels for a selected set of task domains. The research would specify costing considerations, interdependencies between dimensions, relevant task domains, and preliminary descriptions of the situations that would require high levels of performance on each dimension.

The second level of research (Develop Taxonomy) would organize the resulting list of fidelity dimensions into to a taxonomy. The taxonomy could provide the framework for choosing the appropriate fidelity dimensions for any application. The methods that would be used to make the choices are covered in the next level of research.

The third level of research (Dimension Selection) would develop the methods for selecting the appropriate range of fidelity dimensions and levels for any application. The result of this research would be a set of rules that would specify the appropriate fidelity dimensions and levels as a function of the task domain. The rules would be applicable in the set of task domains that were investigated in the first level of research for this topic.

The fourth and final level of this research topic (Analyze for New Domains) would generalize the results of the previous levels to the universe of task domains. This research would repeat the research conducted in the previous levels over a broad range of training domains. We expect that this research would benefit greatly from the results of the previous levels. The results would be a comprehensive list of fidelity dimensions and levels and general procedures for selecting the appropriate range of dimensions for any particular application of the OSBATS model.

Topic 4: Task training hours. The hours of training required to meet the performance standard is a relatively simple judgment for a subject matter expert to make for a task trained in an existing training course. It would also be a relatively easy to extrapolate from similar tasks performed on different weapon systems. The goal of this research topic (New Task Prediction) would be to develop procedures for predicting the training hours required for tasks that currently are not trained. This kind of problem would occur when new capabilities (such as new sensors, weapons, test equipment, and so forth) are developed for a system. In order to apply the OSBATS model to these new tasks, it would be necessary to obtain an estimate of how difficult they are to train on actual equipment.

Topic 5: Task characteristics/instructional features. Currently, the relationship between task characteristics and recommended instructional features is represented by a small set of production rules. These rules were developed for the specific example problem (advanced rotary-wing operations). There is a considerable need to expand both the scope of coverage of the instructional features rules and their level of detail.

The first level of research (Establish Framework) would develop the framework for the instructional feature rule base. This framework will provide more general specifications for the instructional features addressed, the relevant task characteristics, and so forth.

The second level of research (Specific Knowledge Engineering) would involve knowledge engineering activities that would fit into the previously developed framework. In this level, the literature and relevant experts would be consulted to develop specific rules for determining the relevance of instructional features in many task domains.

Topic 6: Fidelity dimensions. The research in Topic 3 would generate a set of fidelity dimensions appropriate to a wide variety of tasks. This topic assumes that the first level of research (generation of cue and response dimensions) for Topic 3 has been conducted. The goal of this topic is to enhance the current fidelity rule base to encompass the increased set of fidelity dimensions and levels.

The first level of research (Characterize Known Relations) examines current sources of information to infer relationships that are already known. This research will examine the psychological literature, rationale for training device designs, and other sources of information relevant to the determination of task cue and response requirements. In addition, the research will involve interviews with experts in the device-design process.

We anticipate that there is considerable information that will be obtained through the first level of research. However, certain critical information will remain unknown. The second step in this topic (Conduct Research Program) would be to conduct a research program to uncover empirically the determiners of task cue and response fidelity requirements.

Topic 7: Instructional features. The current OSBATS model represents instructional features as either present or absent; there is no provision for different levels of instructional features. However, analysis of the model indicates the utility of levels of instructional features, analogous to the levels of fidelity dimensions that are currently in the OSBATS model. The goal of this research topic is to revise the instructional feature selection procedure to accommodate levels of instructional features.

The minimal level of research (Develop Taxonomy) would develop a taxonomy of instructional feature dimensions and add levels. This taxonomy would provide the framework for the analysis methods to be developed in later levels.

The second level (Feature Selection Methods) would revise the current analytical methods used to select instructional features to apply to the revised framework. The result of this task will be a set of procedures that optimize the selection of instructional feature levels. These procedures will be similar to the procedures currently used in the Fidelity Optimization Module, but they will include methods that are specific to instructional features.

The third and fourth levels of research (Characterize Known Relationships and Conduct Research Program) would develop a rule base that is consistent with the instructional feature taxonomy. The third level would characterize known relationships from the original instructional feature rule base, the research literature, and subject matter experts. The fourth level would conduct a program of research specifically designed to uncover critical relationships relevant to the selection of instructional features.

Topic 8: Learning assumptions. The calculations of the OSBATS model are based on several assumptions, some of which have not been validated. This research topic is concerned with testing some of these specific assumptions to determine their validity.

The first level of research (Test Current Assumptions) would involve performing tests on specific assumptions of the OSBATS model, such as the shape of the learning curve and the form of the transfer function. The results of this research could be

used to modify the current assumptions, if necessary, or to estimate the variance of OSBATS predictions.

The second and third level of research (Task Type Framework and Task Type Research) would examine the possibility that different types of tasks (e.g., cognitive, psychomotor, procedural, etc.) are learned differently in a way that can be capitalized upon to improve the predictions of the OSBATS model. The second level of research would develop a task taxonomy to provide a framework for conducting the research and for generating hypotheses for later testing. The third level of research requires designing and conducting empirical research to test the hypotheses.

The final level of research in this topic (Instructional Features and Fidelity Effects) is concerned with the specific assumption of the OSBATS model that fidelity features affect transfer of training as represented by the asymptote of the learning/transfer function, while instructional features affect training efficiency as represented by the time multiplier of the learning/transfer function. This level would design and conduct research to test this assumption.

Topic 9: Aptitude mixture. The current OSBATS model makes its recommendations based on point estimates of learning and transfer parameters. One of the factors that affects the values of these parameters is the aptitude of the students. The goal of this topic is to develop methods that take into account the variance in aptitudes present in the student population in making the model recommendations.

The first level of research (Characterize Relationships) would describe the aptitude relationships that should be accounted for in the model. The effects of aptitude on the training system would be defined and specified.

The second level of research (Model Concept Development) would develop a concept demonstration that would illustrate how aptitude effects would be integrated into the OSBATS model.

Topic 10: Instructional features/fidelity combination. The OSBATS model currently combines its recommendations regarding instructional features and fidelity features in the Fidelity Optimization Module. These recommendations are accomplished by giving instructional features a weight that reflects their importance relative to fidelity features. This weight is currently based on cost comparisons. The goal of this research would be to develop better justified procedures to specify this weighting.

The first level of research (Analytical Evaluation) would require an analytical study that investigates the relative impact

of instructional feature improvements and fidelity improvements on the overall shape of the learning curve. This study could give the rationale for the weight assignment.

The second level (Develop Concept Demonstration) would then develop a concept demonstration of a new OSBATS model (or a revised version of the Fidelity Optimization Module) that would combine instructional feature and fidelity recommendations to determine an overall recommendation regarding the optimal training system designs.

Topic 11: Model advisor. Because of the complexity of the OSBATS model, it is often difficult for the user to fully comprehend the implications of the results. The goal of this research topic is to develop a capability to provide on-line explanations of results and guidance on other analyses that could be performed.

The first step (Explain Results) in this effort would be to develop methods to explain existing results. One of the benefits of this effort is that it would force the system developer to understand all the implications of the model procedures and recommendations. This knowledge could produce new insights and improvements of the basic model itself. The explanatory capability could be implemented as an automated model advisor. This would be at a level considerably above the normal help or explanation screens found in most programs.

The second step (Boundary Cases) would develop procedures that automatically tested boundary cases in order to provide an indication of the robustness of the results for the particular situation. The analyses and interpretive capabilities developed at this level would be relatively fixed and inflexible.

The final step (Suggest Sensitivity Analyses) would develop procedures to suggest, design, and carry out sensitivity analyses. The details of these analyses could be used to explain to the user the reasons for the recommendations of the model.

Topic 12: Prerequisite skills. The OSBATS model currently considers a task as a unitary concept. It does not distinguish the student who doesn't know either the task or its prerequisite skills from the student who doesn't know the task but possesses the prerequisite skills. This distinction may have some impact on the recommendations of the model. The single effort that is considered in this plan would be to conduct an analysis to determine the benefits that would be obtained from considering prerequisite skills in the OSBATS analysis. Further research in this area would be contingent upon the results of the analysis.

Evaluation of Research Issues

This subsection describes both the methods used to conduct the analysis and the results of the analysis.

Analysis Methodology

Our approach to resource allocation is based on a variable's cost and benefit relative to other variables (each research topic is a variable in this case). Our model considers many research topics, and the exploration of each topic will improve the precision of various model components. Exploration of each topic, however, draws from the limited resource of funding. In general, greater exploration of a research topic leads to a better overall model, but requires greater use of the limited resource. The number of potential research plans is the number of combinations of the levels of all 12 research topics and is very large, making it impossible for the unaided designer to select the optimal research plan. The goal of the Resource-Allocation (RA) methodology is to identify the research plan that leads to the greatest benefit for the model with the least resource expenditure. This methodology has a history in training applications (Donnell, Adelman, & Patterson, 1980; Patterson & Adelman, 1981), in the allocation of aircraft to targets (Sticha, Patterson, & Weiss, 1982), and in a number of problems in system design (e.g., Sticha & Patterson, 1981). We conducted the analysis using the EQUITY software package developed by the Decision Analysis Unit at The London School of Economics.

The goal of the RA methodology, stated another way, is to aid the decision maker in determining the appropriate level at which each research topic should be explored. The initial step in determining a research plan is to develop specific research proposals that address a research topic at several levels -- that is, that provide answers of varying completeness. The proposals we developed were described in the previous section. The highest-level proposal provides a reasonably complete exploration of a research topic. The zero-level for a research proposal recommends that no effort be expended on the research topic, and has been left out for clarity. Next, a cost and benefit are assigned to each proposal under each research topic. The benefit values lie in the 0 to 100 range. In most cases, a research proposal at one level depends on completing proposals at lower levels (e.g., the third level is dependent on results from the second level). In these cases, costs and benefits must include the lower-level work.

The next step is to assign importance weights, ranging from 0 to 1000, to each research topic. This assessment reflects the value of that topic to the overall model. As shown in Table 6, we gave research topic 3 the highest weight and topic 12 the lowest weight. For the current research plan the authors

estimated both the cost and benefit values within a research topic and the importance weights of topics. These values were guided by our experience with the OSBATS model, the evaluations by potential users, and sensitivity analyses.

The proposals, assigned cost and benefit values and weighted by the topic importance weights, are entered into the software package to create an RA solution space. This two-dimensional space shows the range of cost-benefit functions for the research plan. The upper bound of the space represents proposal packages that produce the highest benefit at each level of funding. The optimal proposal package can be approximated by locating the point on the upper bound that corresponds to the user's available resources (funding). This point, however, may not represent an actual package -- actual proposal packages may cost less or more than that value. The user must select a nearby point on the curve that represents an actual package. Finally, a sensitivity analysis can be used to determine how manipulations of importance weights and benefit values will affect the resultant solution space.

Assessed Costs and Benefits

Table 6 shows three values in addition to the research proposals. An importance weight is given to each research topic, and cost and benefit values are provided for each research proposal. The procedures taken to arrive at these figures are discussed below.

Importance weights. The critical dimension for importance level was the degree to which the model would be improved if the question posed by the topic were answered. Three of the authors assessed this by first placing each of the 12 research topics into one of three importance levels: high, medium, and low. Next, the group, through discussion, assigned a value between 0 and 1000 to each topic in a category, beginning with those in the high category and ending with the lows. The weights were placed on a ratio scale, with the weight of the most important topic equal to 1000 and the weights for all other topics scaled proportionately across categories. Several informal tests were applied to ensure that the scale had ratio properties:

1. Weights are additive. Thus, if three research topics have weights 800, 400, and 400, the second and third topics together should be evaluated as having the same importance as the first topic alone.
2. Weights are ratio measures. Therefore, a topic with a weight of 800 is evaluated as having twice the importance of a topic with a weight of 400.

Table 6

Cost, Benefit, and Importance Weights of Research Proposals

	Cost	Benefit
VARIABLE 1: Performance Measure Method		CRITERION WTS: 200
1 Performance Measure Summary	200	20
2 Norm-Referenced Measurement	600	60
3 Criterion-Referenced Measurement	900	100
VARIABLE 2: Task Evaluation Factor		CRITERION WTS: 250
1 Compile Dictionary	150	75
2 Develop Analytic Methodology	225	100
VARIABLE 3: Task Cue/Response Requirements		CRITERION WTS: 1000
1 Cue/Response Generation	300	30
2 Develop Taxonomy	500	60
3 Dimension-Selection Method	800	70
4 Analysis for New Domains	1200	100
VARIABLE 4: Task Training Hours		CRITERION WTS: 200
1 New Task Prediction	200	100
VARIABLE 5: Task Characteristics/ Instructional Features		CRITERION WTS: 450
1 Establish Framework	150	30
2 Special Knowledge Engineering	450	100
VARIABLE 6: Fidelity Dimensions		CRITERION WTS: 950
1 Characterize Known Relations	550	70
2 Conduct Research Program	950	100

Table 6 (continued)

Cost, Benefit, and Importance Weights of Research Proposals

		Cost	Benefit
VARIABLE 7: Instructional Features		CRITERION WTS:	750
1 Develop Taxonomy	200	30	
2 Feature-Selection Method	350	45	
3 Characterize Known Relations	700	70	
4 Conduct Research Program	1200	100	
VARIABLE 8: Learning Assumptions		CRITERION WTS:	300
1 Test Current Assumptions	250	30	
2 Task Type Framework	325	50	
3 Task Type Research	625	80	
4 IF/Fidelity Effects	925	100	
VARIABLE 9: Aptitude Mixture		CRITERION WTS:	650
1 Characterize Relations	350	80	
2 Model Concept Demonstration	600	100	
VARIABLE 10: IF/Fidelity Combination		CRITERION WTS:	700
1 Analytical Evaluation	150	80	
2 Develop Concept Demonstration	250	100	
VARIABLE 11: Model Advisor		CRITERION WTS:	750
1 Explain Results	500	50	
2 Boundary Cases	800	70	
3 Suggest Sensitivity	1400	100	
VARIABLE 12: Prerequisite Skills		CRITERION WTS:	100
1 Analyze Relevance	100	100	

Benefit scores. Benefit scores assess the relative contribution of each level within a research topic; they are measured on an interval scale. The benefit for the lowest-level proposal within each topic, which was "no effort," was set to 0. The benefit for the highest-level proposal was set at 100. Again, the authors collectively assigned benefit scores to each research proposal, trying to abide by an interval scale.

Overall benefit values are determined by multiplying the topic importance weights by the benefit score of the specific research proposals. For example, if two research topics have importance weights of 1000 and 700, a proposal under the first topic having the benefit score of 70 would have a resulting value of 70,000, the same value as that for the highest-level proposal (100) under the second topic.

Cost. Obviously, assessed cost should include all costs of conducting the research: labor, subjects, SMEs, materials, planning, execution, analysis, etc. The panel first estimated the time, in months, required to complete each research proposal, and then we applied a simple (i.e., not empirically determined) rule to map time into cost. The assumption behind the rule was that one month of research effort was equivalent to \$10,000 of cost. This simple rule allows us to carry out the resource-allocation example.

Results

The importance weights and benefit and cost values shown in Table 6 were analyzed to select research proposals that would maximize benefit-to-cost ratio. As described above, benefit scores were multiplied by importance weights to determine the total benefit of each research proposal. There were 30 proposals developed from the 12 research topics. Table 7 shows the ordering of these proposals, from highest to lowest, by their benefit-to-cost ratio. Actually, only 24 of the 30 are listed; when a proposal within a research topic had a lower benefit-to-cost ratio than the next-highest proposal under that topic, it was eliminated from the list. In these cases, one can obtain proportionately greater benefit for equivalent cost by selecting the higher-level proposal. Thus, our model does not recommend the less beneficial proposal at any cost.

By addressing items from the top of the list first in the research agenda, one develops the most cost-effective "package" of research proposals. Thus, our proposed research package is driven strongly by the listing in Table 7. Also shown in this Table are cumulative cost and cumulative benefit as proposals are added to an overall research package. Notice that in general, proposals lower in the list include items higher in the list. For example, the third item in the list, the third level for research topic 10, includes the cost and benefit of the first

Table 7

Optimal Order of Inclusion of Research Proposals into the Research Plan

#	VARIABLE	ORDER OF LEVEL	BEST PACKAGES	CUM COST/COST/BENEFIT	
				CUM COST	CUM COST/BENEFIT
1	10 IF/Fidelity Combo	2	Analytical Evaluatn	150	150
2	9 Aptitude Mixture	2	Charact Relations	350	171
3	10 IF/Fidelity Combo	3	Develop Concept Demo	100	194
4	2 Task Evaluatn Factor	2	Compile Dictionary	150	223
5	6 Fidelity Dimensions	2	Charact Known Reltns	550	329
6	3 Task Cue/Resp Rqmts	3	Develop Taxonomy	500	424
7	7 Instructional Feats	2	Develop Taxonomy	200	460
8	4 Task Training Hours	2	New Task Prediction	200	492
9	5 Task Char/Inst Feats	3	Spec Knowledg Engnrng	450	563
10	12 Prerequisite Skills	2	Analyze Relevance	100	579
11	2 Task Evaluatn Factor	3	Develop Analyt Meth	75	589
12	7 Instructional Feats	3	Feature Select Meth	150	607
13	11 Model Advisor	2	Explain Results	500	666
14	6 Fidelity Dimensions	3	Conduct Res Program	400	712
15	3 Task Cue, Resp Rqmts	5	Anal for New Domains	700	775
16	7 Instructional Feats	4	Charac Known Relatns	350	805
17	9 Aptitude Mixture	3	Model Concept Demo	250	825
18	11 Model Advisor	3	Boundary Cases	300	849
19	8 Learning Assumptions	3	Task Type Framework	325	873
20	7 Instructional Feats	5	Conduct Research Pgm	500	909
21	11 Model Advisor	4	Suggest Sen Analysis	600	944
22	8 Learning Assumptions	4	Task Type Research	300	959
23	1 Perf Measure Method	4	Criterion Ref Msrmnt	900	990
24	8 Learning Assumptions	5	IF/Fidelity Effects	300	1000

item, Topic 10's second level. When the third level is added to the proposal package, it replaces the second level. The cost and benefit estimates reflect this fact.

According to the analysis, the most important research proposals are those at the top of the list. An examination of these should provide face validity for our analysis. If the face validity is low, we need to question the assumptions made and the benefit and cost weights assigned to proposals. The first four proposals on the list deal with well defined problems that can be solved with a relatively small effort, and that have a definite impact on the model. These proposals moved to the top of the list primarily because of their low cost, although they also had a moderate benefit.

The next three proposals address the most critical research topics as reflected in their importance weights, Fidelity Dimensions, Task Cue and Response Requirements, and Instructional Features. The cost of these three proposals is considerably greater than the cost of the first four proposals (\$1250K vs. \$750K); the benefit is only slightly greater for the three proposals (237 vs. 223). Thus, even though the issues regarding fidelity and instructional features are critical, they appear on the list after the more specific issues because of their higher cost.

Figure 6 shows a plot of the cumulative benefit and cost of the optimal packages of proposals listed in Table 7. Because the research proposals are chosen according to decreasing incremental benefit-to-cost ratio the optimal points will always lie on a convex curve. This curve represents the upper bound of the solution space; the lower bound is not shown. To select the most cost-effective package, one would simply progress up the curve, including all proposals that could be funded. For example, if three-million dollars (actually \$2,975,000) were available, one could fund the first 12 proposals. Because, some topics are represented by several levels in the group of 12, the lower levels of these topics are removed from the actual package.

The package determined to be optimal is listed in Table 8. Note that the recommendation is made to do no work on three of the 12 research topics. Two of these topics, performance measurement methods and learning assumptions, have relatively low importance weights (less than 5% of the total); the low priority placed on addressing them, therefore, seems justified. Topic 11, model advisor, has an importance weight of slightly less than 12% of the total, making it the third most important topic and one that should warrant high priority. However, the minimal level of research for this topic involved an expense of \$500K, greater than the cost for the lowest level of research for all but one other research topic. Thus, it is not cost-effective to begin

work on the model advisor unless other topics with more immediate payoff are addressed.

In summary, the analysis identified several research topics for which moderate expenditures can have a relatively high payoff. These proposals include the following topics.

1. Develop procedures to estimate the relative importance of fidelity and instructional features in a training device design, and implement these procedures in a concept demonstration.
2. Characterize the relationships by which the range of student aptitude impacts the decisions addressed by the OSBATS model.
3. Compile a comprehensive dictionary of task evaluation factors that provide potential benefits for device-based training.

In addition, the analysis identified three critical topics that require substantial effort, but have the potential for large payoffs to improve the training system design process. These topics involve fidelity dimensions (Topic 6), task cue and response requirements (Topic 3), and instructional features (Topic 7). The ultimate determination of which research topics should be addressed, and the levels at which they should be addressed will depend on the overall research budget and the time period over which the benefits of the recommended research is anticipated. This plan should be reviewed and revised to reflect the results of relevant research as it is conducted.

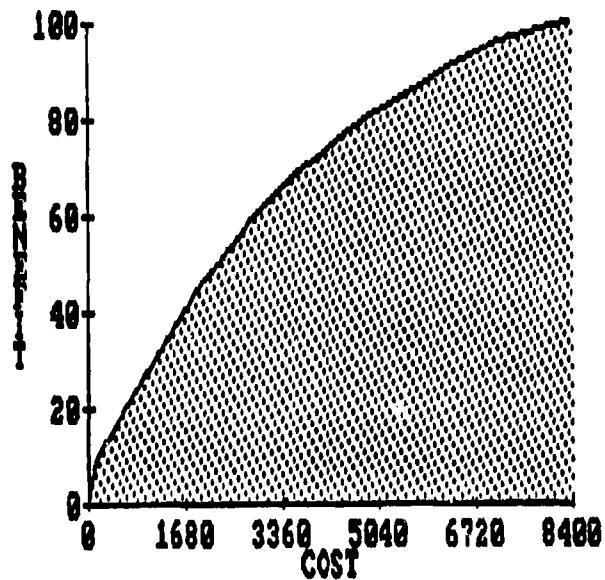


Figure 6. Plot of the benefit and cost of optimal packages of research efforts.

Table 8

Optimal Selection of Research Proposals at a Cost of \$3 Million

TOPIC	COST	WTS	BEN	DESCRIPTION	LEVEL
1 Perf Measure Method	0	32	0	None	0
2 Task Evaluatn Factor	225	40	40	Develop Analyt Meth	2
3 Task Cue/Resp Rqmts	500	159	95	Develop Taxonomy	2
4 Task Training Hours	200	32	32	New Task Prediction	1
5 Task Char/Inst Feats	450	71	71	Spec Knowldg Engnrgn	2
6 Fidelity Dimensions	550	151	106	Charact Known Reltns	1
7 Instructional Feats	350	119	54	Feature Select Meth	2
8 Learning Assumptions	0	48	0	None	0
9 Aptitude Mixture	350	103	83	Charact Relations	1
10 IF/Fidelity Combo	250	111	111	Develop Concept Demo	2
11 Model Advisor	0	119	0	None	0
12 Prerequisite Skills	100	16	16	Analyze Relevance	1
	2975		607		

REFERENCES

Adams, J. (1957). Some considerations in the design and use of dynamic flight simulators (Tech. Note 57-51). Lackland AFB, TX: Air Force Personnel and Training Research Center.

Adams, J. A. (1979). On the evaluation of training devices. Human Factors, 21, 711-720.

Allbee, K. E., & Sample, C. A. (1981). Aircrew training devices: Life-cycle cost and worth of ownership (AFHRL-TR-80-34). Brooks AFB, TX: Air Force Human Resources Laboratory.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Armstrong, J. S. (1985). Long-range forecasting: From crystal ball to computer (2nd ed.). New York: John Wiley.

Armstrong, J. S. (1986). The ombudsman: Research on forecasting: A quarter-century review, 1960-1984. Interfaces, 16, 89-109.

Atkinson, R. C., Bower, G. H., & Crothers, E. J. (1965). An introduction to mathematical learning theory. New York: John Wiley.

Atkinson, R. C., & Crothers, E. J. (1964). A comparison of paired-associate learning models having different acquisition and retention axioms. Journal of Mathematical Psychology, 1, 285-315.

Bailey, J., Hughes, R. G., & Jones, W. E. (1980). Application of backward chaining to air-to-surface weapons delivery training (AFHRL Tech. Rep. 79-63). Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory. (AD-A085 610)

Baum, D. R., Smith, D. A., Hirshfield, S. F., Klein, G. A., Swezey, R. W., & Hays, R. T. (1982). Specification of training simulator fidelity: A research plan (Tech. Rep. 558). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A127 346)

Bickley, W. R. (1980). Training device effectiveness: formulation and evaluation of a methodology (Research Rep. 1291). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A122 777)

Binford, J. R., & Gettys, C. (1965). Nonstationarity in paired-associate learning as indicated by a second-guess procedure. Journal of Mathematical Psychology, 2, 190-195.

Blaiwes, A. S., Puig, J. A., & Regan, J. J. (1973). Transfer of training and the measurement of training effectiveness. Human Factors, 15, 523-533.

Block, J. H., & Burns, R. B. (1976). Mastery Learning. In L. Shulman (Ed.), Review of research in education. Itasca, IL: F. E. Peacock.

Boehm, B. (1981). Software Engineering Economics. Englewood Cliffs, NJ: Prentice-Hall.

Bower, G. H. (1961). Application of a model to paired-associate learning. Psychometrika, 26, 255-280.

Bower, G. H. (1967). A Multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory (Vol. 1). New York: Academic Press.

Bower, G. H., & Theios, J. (1964). A learning model for discrete performance levels. In R. C. Atkinson (Ed.), Studies in mathematical psychology. CA: Stanford University Press.

Braby, R., Henry, J. M., Parrish, W. F., Jr., & Swope, W. M. (1975). A technique for choosing cost-effective instructional delivery systems (TAEG Rep. No. 16). Orlando, FL: Department of the Army, Training Analysis and Evaluation Group.

Brainerd, C. J., Howe, M. L., & Desrochers, A. (1982). The general theory of two-stage learning: A mathematical review with illustrations from memory development. Psychological Bulletin, 91(3), 634-665.

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J. P., King, F. J., & Hannum, W. J. (1975). Interservice procedures for Instructional Systems Development (TRADOC PAM 350-30). Fort Monroe, VA: U.S. Army Training and Doctrine Command.

Briston, C. A., & Burger, W. J. (1976). Transfer of training effectiveness: A7E night carrier landing trainer (NCLT) device 2F103 (Tech. Rep. NAVTRAEEQUIPCEN 74-C-00790-9-1). Orlando, FL: Naval Training Equipment Center. (AD A028-836)

Briggs, G. E. (1961). On the scheduling of training conditions for the acquisition and transfer of perceptual motor skills (NAVTRADEVVCEN Rep. No. 836-1). Orlando, FL: Naval Training Equipment Center.

Bush, R., Jr., & Mosteller, F. (1955). Stochastic models for learning. New York: John Wiley.

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human-computer interaction. Hillsdale, NJ: Erlbaum.

Carter, G., & Trollip, S. (1980). A constrained maximization extension to incremental transfer effectiveness, or, how to mix your training technologies. Human Factors, 22(2), 141-152.

Caro, P. W., Jr., (1972). Transfer of instrument training and the synthetic flight training system (NAVTRADEVCECEN IH-206). In Proceedings of the 5th Naval Training Device Center/Industry Conference. Orlando, FL: Naval Training Device Center.

Caro, P. W., Jr. (1973). Aircraft simulators and pilot training. Human Factors, 15, 503-510.

Caro, P. W., Pohlman, L. D., & Isley, R. N. (1979). Development of simulator instructional feature design guides (TR 79-12). Pensacola, FL: Seville Research Corporation.

Chant, V. G., & Atkinson, R. C. (1973). Optimal allocation of instructional effort to interrelated learning strands. Journal of Mathematical Psychology, 10, 1-25.

Condon, C. F., Ames, L. L., Hennessey, J. R., Shriner, E. L., & Seeman, R. E. (1974). Flight simulator maintenance training: Potential use of state-of-the-art simulation techniques (AFHRL Tech. Rep. 79-19). Brooks AFB, TX: Human Resources Laboratory.

Cronholm, J. N. (1985). The optimization of training systems. Proceedings of Symposium on the Transfer of Training to Military Operational Systems (pp. 237-259). Brussels, Belgium: NATO.

Cross, K. D., & Gainer, C. A. (1987). An enumeration of research to determine the optimal design and use of Army flight training simulators (Technical Report 763). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A191 242)

Dawdy, E. D., Chapman, W. A., & Frederickson, E. W. (1981a). REMBASS final report, Vol. 1 Executive summary (Contract No. DABT60-80-C-0056). Ft. Huachuca, AZ: Applied Science Associates.

Dawdy, E. D., Chapman, W. A., & Frederickson, E. W. (1981b). REMBASS final report, Vol. 2 (Contract No. DABT60-80-C-0056). Ft. Huachuca, AZ: Applied Science Associates.

Dohme, J. A., & Millard, S. L. (In preparation). Training helicopter initial entry students in simulation (THIESIS). Fort Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Donnell, M. L., Adelman, L., & Patterson, J. F. (1980). Development of a computerized training requirements and cost evaluation system for the U.S. Marine Corps (Final Technical Report TR 80-5-315.3). McLean, VA: Decisions and Designs, Inc.

Dougherty, D. J., Houston, R. C., & Nicklas, D. R. (1957). Transfer of training in flight procedures from selected ground training devices to the aircraft (Human Engineering Tech. Rep. NAVTRADEVcen 71-16-16). Port Washington, NY: U.S. Naval Training Device Center.

Estes, W. K. (1959). The statistical approach to learning theory. In S. Koch (Ed.), Psychology: A study of a science (Vol. 2). New York: McGraw-Hill.

Estes, W. K. (1960). Learning theory and the new "mental chemistry." Psychological Review, 67, 207-223.

Evans, R. M., Scott, P. G., & Pfeiffer, M. G. (1984). SH-3 helicopter flight training: An evaluation of visual and motion simulation in device 2F64C (TAEG Tech. Rep. 161). Orlando, FL: Naval Training Equipment Center.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning. New York: Academic Press.

Fitts, P. M., Bahrick, H. P., Noble, M. E., & Briggs, G. E. (1961). Skilled performance. New York: John Wiley.

Fitts, P. M., & Posner, M. I. (1967). Human performance. Belmont, CA: Brooks/Cole Publishing Co.

Fletcher, J. D. (1985). Some approaches to model-based instruction (Final Report). Eugene, OR: Center for Advanced Technology in Education, University of Oregon.

Flexman, R. E., Roscoe, S. N., Williams, A. C., & Williges, B. H. (1972). Studies in pilot training: The anatomy of transfer. Aviation Research Monographs, 2, (1). Urbana-Champaign: Aviation Research Laboratory, Institute of Aviation, University of Illinois.

Freida, J. S. (1983). Approach to fidelity in training simulation (Research Note 83-03). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A125 376)

Frederickson, E. W., Hawley, J. K., & Whitmore, P. G. (1983). The training developer's decision aid: A computer-aided approach to training design (Contract No. MDA903-80-C-0160). El Paso, TX: Applied Science Associates.

Gagne, R. M., (1954). Training devices and simulators: Some research issues. American Psychologist, 9, 95-107.

Gagne, R. M. (1968). A systems approach to adult learning (Tech. Rep. No. 15, 1-14). Washington University Department of Psychology.

Gagne, R. M. (1973). Learning and instructional sequence. In F. N. Kerlinger (Ed.), Review of research in education (pp. 3-33). Itasca, IL: F.E. Peacock.

Gagne, R. M., & Briggs, L. J. (1979). Principles of instructional design (2nd ed.). New York: Holt, Rinehart and Winston.

Gagne, R. M., Reiser, R. A., & Larsen, J. (1981). A learning based model for media selection: Description (Research Product 81-25A). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A109 472)

Goldberg, I. (1988). Training effectiveness and cost iterative technique (TECIT). Vol. I: Training effectiveness analysis (Research Note 88-35). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A196 727)

Goldberg, S. L., Drilling, M., & Dressell, J. D. (1981). Mastery training: Effect on skill retention (Technical Report 513). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A120 762)

Golub, G. H., & Pereyra, V. (1972). The differentiation of pseudo-inverses and non-linear least squares whose variables separate (Tech. Rep. No. 261). Palo Alto, CA: Computer Science Department, Stanford University.

Gray, T., & Fuller, R. (1977). Effects of simulator training and platform motion on air-to-air surface weapons delivery training (Tech. Rep. 77-29). Brooks AFB, TX: U.S. Air Force Human Resources Laboratory.

Greene, J. G. (1967). Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. Journal of Mathematical Psychology, 4, 430-472.

Grice, G. R. (1968). Stimulus intensity and response evocation. Psychological Review, 75(5), 359-373.

Hagin, W. V. (1976). Platform motion in flight simulators: Critical or nice? In Proceedings of the Society for Applied Learning Technology. Washington, DC.

Hagman, J. D. (1980a). Effects of presentation- and test-trial training on motor acquisition and retention (Tech. Rep. 431). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A086 406)

Hagman, J. D. (1980b). Effects of training schedule and equipment variety on retention and transfer of maintenance skill (Research Report 1309). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A120 167)

Hagman and Rose (1983). Retention of military tasks: A review. Human Factors, 25(2), 199-123.

Harris, J. H., Ford, P., Tufano, D., & Wiggs, C. (1983). Application of transfer forecast methods to armor training devices (HumRRO FR-TRD(VA)-83-3). Alexandria, VA: Human Resources Research Organization.

Hawley, J. K., & Dawdy, E. D. (1981). Cost and information effectiveness analysis (CIEA): A methodology for evaluating device operational readiness assessment capability (Final Rep.). Alexandria, VA: Applied Science Associates.

Hawley, J. K., & Frederickson, E. W. (1983). The training development decision support system: An aid to the design of cost-effective training programs (397-401). In A. T. Pope & L. D. Haugh (Eds.), Proceedings of the Human Factors Society - 27th Annual Meeting, Vol. 1. Norfolk, VA.

Hayes, R. T. (1980). Simulator fidelity: A concept paper (Technical Report 490). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A101 475)

Heffley, R. K., Clement, W. F., Ringland, R. F., Jewell, W. F., Jex, H. R., McRuer, D. T., & Carter, V. E. (1981). Determination of motion and visual system requirements for flight training simulators (Tech. Rep. 546). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A117 555)

Hogarth, R. M. (1980). Judgement and choice: The psychology of decisions. New York: John Wiley.

Holman, G. L. (1979). Training effectiveness of the CH-47 flight simulator (Research Report 1209). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A072 317)

Holmgren, J. E., Hilligoss, R. E., Swezey, R. W., & Eakins, R. C. (1979). Training effectiveness and retention of training extension course (TEC) instruction in the combat arms (Research Report 1208). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A069 942)

Hughes, R. G. (1979). Advanced training features: Bridging the gap between in-flight and simulator-based models of flying training (AFHRL Tech. Rep. 78-96). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hughes, R., Hannan, S., & Jones, W. (1979). Application of flight simulator record/playback feature (AFHRL Tech. Rep. 79-52). Brooks AFB, TX: Air Force Human Resources Laboratory.

Hughes, R. G., Lintern, G., Wightman, D. C., & Brooks, R. B. (1981). On the use of a flight simulator's freeze feature during acquisition of a carrier landing task. Proceedings of the Third Interservice/Industry Training Equipment Conference and Exhibition, 1, 149-157.

Hull, C. L. (1943). Principles of behavior. New York: Appleton-Century-Crofts.

Hull, C. L. (1952). A behavior system: An introduction to behavior theory concerning the individual organism. New Haven, CT: Yale University Press.

Isley, R., & Miller, E. (1976). The role of automated training in future Army flight simulators (FR-76-27). Alexandria, VA: Human Resources Research Organization.

Jacobs, R. S., & Roscoe, S. N. (1975). Simulator cockpit motion and the transfer of initial flight training. In Proceedings of the Human Factors Society 19th Annual Meeting. Santa Monica, CA: Human Factors Society.

Jaeger, R. J., Agarwal, G. C., & Gottlieb, G. L. (1980). Predictor operator in pursuit and compensatory tracking. Human Factors, 22(4), 497-506.

Jones, E. R., Hennessy, R. T., & Deutsch, S. (Eds.). (1985). Human factors aspects of simulation: Report of the working group on simulation (Contract No. N00014-85-C-0093). Washington, DC: Committee on Human Factors, National Research Council.

Jorgensen, C. C. (1984). Nuclear power plant personnel qualifications and training: TSORT—an automated technique to assign tasks to training strategies. Oak Ridge, TN: Oak Ridge National Laboratory.

Kieras, D. E. (1985). The role of prior knowledge in operating equipment from written instructions (Tech. Rep. No. 19). Ann Arbor: University of Michigan.

Kinkade, R. G., & Wheaton, G. R. (1972). Training device design. In H. P. Van Cott, & R. G. Kinkade (Eds.), Human engineering guide to equipment design (Revised Ed.). Washington, DC: U.S. Government Printing Office.

Kintsch, W. (1970). Learning, memory, and conceptual processes. New York: John Wiley.

Klein, G. C., Kane, J. J., Chinn, A. A., & Jukes, A. D. (1978). Analyzing training device effectiveness in cases where test data is inconclusive. In Proceedings of the Human Factors Society 22nd Annual Meeting (pp. 152-156). Detroit, MI.

Knapp, M. I., & Orlansky, J. (1983). A cost element structure for defense training (IDA Paper P-1709). Alexandria, VA: Institute for Defense Analyses. (AD-A139-164)

Knerr, C. M., Nadler, L. B., & Dowell, S. K. (1984). Training transfer and effectiveness models (HumRRO Final Report FR-TRD(VA)-84-1). Alexandria, VA: Human Resources Research Organization.

Knerr, C. M., & Sticha, P. J. (1985). Models of learning and performance of armor skills. Proceedings of the Symposium on the Military Value and Cost-Effectiveness of Training (pp. 491-512). Brussels, Belgium: NATO.

Kribs, H. D., Simpson, A. C., & Mark, L. J. (1983). Automated instructional media selection (AIMS) (Tech. Rep. NAVTRAEEQUIPCEN 79-C-0104-1). Orlando, FL: Naval Training Equipment Center.

Lewis, C. H. (1979). Speed and practice. Unpublished manuscript. (Available from C. H. Lewis, IBM Research Center, Yorktown Heights, NY 10598).

Lewis, D., McAllister, D. E., & Bechtoldt, H. D. (1953). Correlational study of performance during successive phases of practice for the standard and reversed tasks on the SAM complex coordinator. Journal of Psychology, 36, 111-126.

Lintern, G. T., & Roscoe, S. N. (1978). Transfer of landing skill after training with supplementary visual cues. In Sixth Symposium on Psychology in the Department of Defense. Colorado Springs, CO: U.S. Air Force Academy.

Lintern, G., Thomley, K. E., Nelson, B. E., & Roscoe, S. N. (1984). Content, variety, and augmentation of simulated visual scenes for teaching air-to-ground attack (Tech. Rep. NAVTRAEEQUIPCEN 81-C-0105-3). Orlando, FL: Naval Training Equipment Center.

Logicon, Inc. (1985). Instructional support feature guidelines (AF Contract No. F33615-84-C-0054). San Diego, CA: Author.

Lorton, P. V. (1973). Doctoral Dissertation. CA: Stanford University.

Marcus, G. H., Patterson, J. T., Bennett, C. D., & Gershan, B. S. (1980). Cost-effectiveness methodology for aircrew training devices: Model development and users' handbook (Final Report AFHRL-TR-79-39). Brooks AFB, TX: Air Force Human Resources Laboratory.

Martin, E. L., & Waag, W. L. (1978a). Contributions of platform motion to training effectiveness: Study II. Aerobatics (Tech. Rep. AFHRL-TR-78-520). Williams AFB, AZ: Air Force Human Resources Laboratory, Flying Training Division. (AD-A064-305)

Martin, E. L., & Waag, W. L. (1978b). Contributions of platform motion to training effectiveness: Study I. Basic contact (Tech. Rep. AFHRL-TR-78-15). Williams AFB, AZ: Air Force Human Resources Laboratory, Flying Training Division. (AD-A058-416)

Martino, J. P. (1983). Technological forecasting for decision making (2nd ed.). New York: North-Holland.

Matheny, W. G. (1974). Training research programs and Plans: Advanced simulation in undergraduate pilot training (ASUPT) (Report LSI-TR-74-2). Phoenix, AZ: Human Resources Laboratory, Flight Training Division.

Matheny, W. G. (1975). Investigation of the performance equivalence method for determining training simulator and training methods requirements. American Institute of Aeronautics and Astronautics 13th Aerospace Sciences Meeting. Pasadena, CA. (AIAA Paper 75-108)

Matlick, R. H., Rosen, M. H., & Berger, D. C. (1980). Cost and training effectiveness analysis performance guide (Final Report MDA 903-79-C-0326). Springfield, VA: Litton Mellonics.

Mazur, J. E., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. Psychological Bulletin, 85(6), 1256-1274.

Miller, R. B. (1954). Psychological considerations in the design of training equipment (WADC Tech. Rep. 54-563). Wright-Patterson AFB, OH: Wright Air Development Center.

Montemerlo, M. D. (1975). Instructional systems development, state-of-the-art and directions for the future. Paper presented at 8th Naval Training Equipment Center, Industrial Conference, Orlando, FL.

Mudd, S. (1968). Assessment of the fidelity of dynamic simulators. Human Factors, 10, 351-358.

Narva, M. A. (1979). Development of a systematic methodology for the application of judgmental data to the assessment of training device concepts (ARI Research Memorandum 79-7). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A077 980)

Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Noble, M., Trumbo, D., Ulrich, L., & Cross, K. (1966). Task predictability and the development of tracking skill under extended practice. Journal of Experimental Psychology, 72(1), 85-94.

Norman, D. A. (1966). Acquisition and retention in short-term memory. Journal of Experimental Psychology, 72, 369-381.

Norman, D. A. (1969). An introduction to human information processing. New York: John Wiley.

Orlansky, J., Knapp, M. I., & String, J. (1984). Operating costs of aircraft and flight simulators (IDA paper P-1733). Alexandria, VA: Institute for Defense Analyses.

Osborn, W. C., Campbell, C. H., & Harris, J. H. (1979). The retention of tank crewman skills (Research Report 1234). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A086 399)

Osgood, C. E. (1949). The similarity paradox in human learning: A resolution. Psychological Review, 56, 132-143.

Parker, J., & Downs, J. (1961). Selection of training media (Tech. Rep. 61-473). Wright-Patterson AFB, OH: Aerospace Medical Laboratory.

Patterson, J. F., & Adelman, L. (1981). Training requirements and cost evaluation system (TRACES) for Marine Air Command and Control: Phase I (Interim Report PR 81-26-329). McLean, VA: Decisions and Designs, Inc.

Paulson, J. A. (1973). An evaluation of instructional strategies in a simple learning situation (Doctoral dissertation and Tech. Rep. No. 209). Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA.

Payne, T. A., Hirsh, D. L., Semple, C. A., Farmer, J. R., Spring, W. G., Sanders, M. S., Wimer, C. A., Carter, V. E., Hu, A. (1976, March). Experiments to evaluate advanced flight simulation in air combat pilot training: Volume 1. Transfer of learning experiment (Contract Number 62269-74-C-0314). Hawthorne, CA: Northrop Corporation, Aircraft Division.

Pew, R. W. (1966). The acquisition of hierarchical control over the temporal organization of a skill. Journal of Experimental Psychology, 71(5), 764-771.

Plaats, J. V., Butler, G., Hays, S., & Atkins, W. R. (1982). Training analysis support computer system (TASCS): Automated user assistance for instructional systems development projects. San Diego, CA: Logicon Inc.

Pohlman, L. D., & Reed, J. C. (1978). Air-to-air combat skills: Contribution of platform motion to initial training (Tech. Rep. AFHRL-TR-78-53). Williams AFB, AZ: Air Force Human Resources Laboratory, Flying Training Division. (AD-A0620738)

Povenmire, H. K., & Roscoe, S. N. (1973). Incremental transfer effectiveness of a ground-based general aviation trainer. Human Factors, 15, 534-542.

Pozella, D. J. (1983). Aircrew training devices: Utility and utilization of advanced instructional features (Phase I--Tactical Air Command) (AFHRL Tech. Rep. 83-22). Williams AFB, TX: Air Force Human Resources Laboratory.

Prophet, W. W., & Boyd, H. A. (1970). Device-task fidelity and transfer of training: Aircraft cockpit procedures training (TR-70-10). Alexandria, VA: Human Resources Research Organization.

Raaijmakers, J. G. W., & Shiffin, R. M. (1981). Search of associative memory. Psychological Review, 88(2), 93-134.

Resnick, L. B., & Wang, M. C. (1969). Approaches to the validation of learning hierarchies. PA: Learning Research and Development Center, University of Pittsburgh.

Restle, F. (1964). Sources of difficulty in learning paired-associates. In R. C. Atkinson (Ed.), Studies in mathematical psychology. CA: Stanford University Press.

Restle, F. (1965). The significance of all-or-none learning. Psychological Bulletin, 69, 329-324.

Restle, F., & Greeno, J. G. (1970). Introduction to mathematical psychology. Reading, MA: Addison-Wesley.

Rigg, K. E., & Gray, B. B. (1980). Estimating skill training and retention functions through instructional model analysis (Contract MDA903-79-C-0278). Monterey, CA: McFann, Gray & Associates, Inc.

Rock, I. (1957). The role of repetition in associative learning. American Journal of Psychology, 70, 186-193.

Roscoe, S. N. (1971). Incremental transfer effectiveness. Human Factors, 15, 534-567.

Roscoe, S. N. (1980). Aviation psychology. Ames: The Iowa State University Press.

Rose, A. M., Czarnolewski, M. Y., Gragg, F. E., Austin, S. H., Ford, P., Doyle, J., & Hagman, J. D. (1985). Acquisition and retention of soldiering skills (Technical Report 671). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A160 336)

Rose, A. M., Radtke, P. H., Shettel, H. H., & Hagman, J. D. (1985). User's manual for predictin military task retention (Revised March 1985). Washington, DC: American Institutes for Research.

Rose, A. M., Wheaton, G. R., & Yates, L. G. (1985). Forecasting device effectiveness: Volume I. Issues (Tech. Rep. 680). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A159 576)

Ryan-Jones, D. L. (1984). Aviator training requirements and simulation fidelity (Research Note). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (Unpublished)

Ryan, L. E., Scott, P. G., & Browning, R. F. (1978). The effects of simulator landing practice and the contribution of motion to P-3 pilot training (TAEG Rep. 63). Orlando, FL: Navy Training Analysis and Evaluation Group.

Schendel, J. D., & Hagman, J. D. (1980). On sustaining procedural skills over prolonged retention intervals (Research Report 1298). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A120 758)

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). Retention of motor skills: A review (Technical Paper 313). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A061 338)

Seiple, C. A., Cotton, J. C., & Sullivan, D. J. (1981). Aircrew training devices: Instructional support features (AFHRL-TR-80-58). Brooks AFB, TX: Air Force Human Resources Laboratory.

Seiple, C. A., Hennessy, R. T., Sanders, M. S., Cross, B. K., Beith, B. H., & McCauley, M. E. (1981). Aircrew training devices: Fidelity features (AFHRL-TR-80-36). Brooks AFB, TX: Air Force Human Resources Laboratory.

Sheppard, D. J., (1985). Visual and part-task manipulations for teaching simulated carrier landings (NAVTRAECIPCEN Tech. Rep. 81-C-0105-9). Orlando, FL: Naval Training Equipment Center.

Smode, A., Gruber, A., & Ely, J. (1963). Human factors technology in the design of simulators for operator training (Tech Rep. 1103-1). Port Washington, NY: U.S. Naval Training Device Center.

Snoody, G. S. (1926). Learning and stability. Journal of Applied Psychology, 10, 1-36.

Spears, W. D. (1985). Measurement of learning and transfer through curve fitting. Human Factors, 27(3), 251-266.

Spence, K. W. (1956). Behavior theory and conditioning. New Haven, CT: Yale University Press.

Stevens, J. C., & Savin, H. B. (1962). On the form of learning curves. Journal of the Experimental Analysis of Behavior, 5(1), 15-18.

Singer, M. J., & Sticha, P. S. (1987). Designing training devices: The optimization of simulation-based training systems. Proceedings of the 9th Interservice/Industry Training Systems Conference (pp. 26-31). Washington, DC: American Defense Preparedness Association.

Sticha, P. J. (1989). Normative and descriptive models for training-system design. In G. R. McMillan (Ed.), Applications of human performance to systems design. New York, NY: Plenum Publishing Corporation.

Sticha, P. S., Blacksten, H. R., & Buedo, D. M. (1986). Optimization of simulation-based training systems. Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics (pp. 1373-1377). New York: The Institute of Electrical and Electronics Engineers.

Sticha, P. S., Blacksten, H. R., Buedo, D. M., Singer, M. J., Gilligan, E. L., Mumaw, R. J., & Morrison, J. E. (1990). Optimization of simulation-based training systems: Model description, implementation, and evaluation (Technical Report 896). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences. (AD A227 249)

Sticha, P. J., Edwards, T. D., & Patterson, J. F. (1984). An analytic model of learning and performance of armor procedures (Research Note 84-24). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A138 000)

Sticha, P. J., & Knerr, C. M. (1984). Task-element and individual differences in procedural learning and retention: A model-based analysis (Research Note 84-1). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A136 789)

Sticha, P. J., & Patterson, J. F. (1981). Evaluation of video teleconferencing systems (Final Report TR 80-8-317). McLean, VA: Decisions and Designs, Inc.

Sticha, P. J., Patterson, J. F., & Weiss, J. J. (1982). Approximate reasoning methods for decision aids (Technical Report TR 82-6-157). McLean, VA: Decisions and Designs, Inc.

String, J., & Orlansky, J. (1977). Cost-effectiveness of flight simulators for military training: Volume II. Estimating costs of training in simulators and aircraft (Final Report DAHC15-73-C-0200). Alexandria, VA: Institute for Defense Analyses.

Suppes, P., Fletcher, J. D., & Zanotti, H. (1976). Models of individual trajectories in computer-assisted instruction for deaf students. Journal of Educational Psychology, 68(2), 117-127.

Swezey, R. W., & Evans, R. A. (1981). Guidebook for users of TRAINVICE II (Research Note 81-29). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A126 197)

Thode, W. F., & Walker, R. A. (1983). A model for comparing training costs in a complex military training system. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal. (ED 232 262)

Thorndike, E. L. (1903). Educational psychology. New York: Lemke and Buechner.

Thorndike, E. L., Cobb, M. V., Orleans, S. S., Symonds, P. M., Wald, E., & Woodyard, E. (1923). The psychology of algebra. New York: Macmillan.

Thorpe, J. A., Varney, N. C., McFadden, R. W., LeMaster, W. D., & Short, L. H. (1978). Training effectiveness of three types of visual systems for KC-135 flight simulators (Tech. Rep. AFHRL-TR-78-16). Williams AFB, AZ: Air Force Human Resources Laboratory, Flying Training Division. (AD-A060-253)

Thurstone, L. L. (1919). The learning curve equation. Psychological Monographs, 26(114), 51.

Thurstone, L. L. (1927). A law of comparative judgement. Psychological Review, 34, 273-286.

Tosti, D. T., & Ball, J. R. (1969). A behavioral approach to instructional design and media selection. Audiovisual Communication Review, 5-25.

Trumbo, D., Noble, M., Cross, K., & Ulrich, L. (1965). Task predictability in the organization, acquisition, and retention of tracking skill. Journal of Experimental Psychology, 70, (3), 252-263.

Tufano, D. R., & Evans, R. A. (1982). The prediction of training device effectiveness: A review of Army models (Technical Report 613). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A146 937)

Underwood, B. J., & Keppel, G. (1962). One-trial learning. Journal of Verbal Learning and Verbal Behavior, 1, 1-13.

U.S. Army TRADOC Systems Analysis Activity (TRASANA) (1977). REDEYE weapons system training effectiveness analysis (TRASANA Tech. Rep. 21-77). White Sands, NM: Author.

U.S. Department of Defense, Defense Science Board. (1982). Summer study on training and training technology. Office of the Under Secretary of Defense Research and Engineering. Washington, DC: Author.

Vineberg, R., & Joyner, J. N. (1980). Instructional system development (ISD) in the armed services: Methodology and application (Tech. Rep. 80-1). Alexandria, VA: Human Resources Research Organization.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., & Holding, D. H. (1976). Evaluation of the effectiveness of training devices: Literature review and preliminary model. (Research Memorandum 76-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A076 809)

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Leonard, R. L., Jr., & Boycan, G. G. (1976). Evaluation of effectiveness of three burst-on-target trainers (ARI Research Memorandum 76-18). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A076 820)

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. Journal of Mathematical Psychology, 2, 418-455.

Wickelgren, W. A. (1974a). Single-trace fragility theory of memory dynamics. Memory and Cognition, 2, 775-780.

Wickelgren, W. A. (1974b). Strength/resistance theory of the dynamics of memory storage. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology, (Vol. 1). San Francisco: W. H. Freeman.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. Journal of Mathematical Psychology, 3, 316-347.

Willis, R. P., Guha, P., & Hunter, D. (1988). The optimization of simulation-based training: Model Data Collection and Utilization. Orlando, FL: U.S. Army Research Institute. (unpublished)

Wood, M. (1977). The fidelity issue in visual simulation. In Proceedings of the 1977 Image Conference. Williams AFB, AZ: U.S. Air Force Human Resources Laboratory.

Woomer, C., & Carico, D. (1977). A program for increased flight fidelity in helicopter simulation. 10th NTEC/Industry Conference Proceedings. Orlando, FL: Naval Training Equipment Center.

Young, P. R., Luster, P. K., Stock, J. A. (PAR Government Systems Corporation), Mumaw, R. S., & Sticha, P. J. (HumRRO) (1986). Optimization of simulation-based training systems. Volume I: Plans for model implementation, evalution and research (HumRRO Final Report FR-PRD-86-13). Alexandria, VA: Human Resources Research Organization.

APPENDIX A. SYNOPSIS OF FLIGHT SIMULATOR TRANSFER-OF-TRAINING STUDIES

REFERENCE: Bickley (1980)

PRIMARY STUDY OBJECTIVE: Evaluate cost and training effectiveness of prototype AH-1 Flight Weapons Simulator (FWS) and develop trade-off functions for use in defining the optimal mix of aircraft and simulator training.

SUBJECT POPULATION: Rated Army helicopter aviators enrolled in the AH-1 Aircraft Qualification Course (AQC); 21 experimental group and 25 control group subjects.

TRANSFER AIRCRAFT

- Type: Army Rotary Wing
- Designation: AH-1 (Cobra)

FLIGHT SIMULATOR CHARACTERISTICS

- Name: AH1FWS
- Motion System: 6df Motion Platform
- Visual System: Camera-Modelboard System; pilot FOV is 36-degree vertical and 101-degree horizontal (two windows); gunner FOV is 36-degree vertical and 48-degree horizontal (one window).

INDEPENDENT VARIABLES INVESTIGATED (if any): Number of practice iterations in flight simulator.

TASKS TRAINED IN FLIGHT SIMULATOR: Thirty-one flight and weapons task; takeoffs, landings, airwork, emergency procedures, weapons procedures/firing, etc.

PERFORMANCE MEASURES USED: Instructor pilot ratings number of practice iterations, and training time spent performing iterations. (Note: Subjects received a prescribed number of practice iterations in the flight simulator and were trained to criterion on the aircraft.)

KEY FINDINGS: Positive training transfer demonstrated for most tasks. Demonstrated viability and utility of the model and methodology.

REFERENCE: Brictson and Burger (1976)

PRIMARY STUDY OBJECTIVE: Assess training effectiveness of the Night Carrier Landing Trainer.

SUBJECT POPULATION: Novice aviators (320-330 jet hrs) and experienced aviators (1140-1290 jet hrs).

TRANSFER AIRCRAFT

- Type: Navy Fixed Wing
- Designation: A-7E

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Night Carrier Landing Trainer
- Motion System: 3 df Motion Platform
- Visual System: Computer Generated Display; FOV is 30-degree vertical and 40-degree horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): Aviator experience level (2 levels).

TASKS TRAINED IN FLIGHT SIMULATOR: Night carrier approaches and landings.

PERFORMANCE MEASURES USED: Radar measures of aircraft variables during final approach; an objective measure derived from wire arrestment or wave-off data and weighted according to quality by LSO consensus; percent of final approaches resulting in touching down beyond arrestment wires; frequency distribution of wire numbers caught during carrier qualification; subjective LSO scores; pilot questionnaires; and the number of aviators who passed/failed carrier qualification.

KEY FINDINGS: Positive transfer was demonstrated only for novice aviators.

REFERENCE: Browning, McDaniel, Scott, and Smode (1982)
McDaniel, Scott, and Browning (1983)
Evans, Scott, and Pfeiffer (1984)

PRIMARY STUDY OBJECTIVE: Series of studies to assess effectiveness of simulator training with (a) both visual system and motion system in operation, (b) only motion system in operation and (c) neither motion or visual system in operation.

SUBJECT POPULATION: Newly designated Naval aviators undergoing replacement pilot training (Helicopter Antisubmarine). Visual and motion group, N=19; motion only group, N=29; no visual or motion, N=11; and fly only group, N=16.

TRANSFER AIRCRAFT

- Type: Navy Rotary Wing
- Designation: SH-3 Sea King

FLIGHT SIMULATOR CHARACTERISTICS

- Name: SH-3FS (Device 2F64C)
- Motion System: 6 df Motion Platform
- Visual System: Computer Generated (VITAL IV, McDonnell Douglas); FOV not stated.

INDEPENDENT VARIABLES INVESTIGATED (if any): Presence or absence of visual system and motion system during simulator training.

TASKS TRAINED IN FLIGHT SIMULATOR: 22 tasks: 5 before takeoff, 13 airwork, and 4 emergency tasks.

PERFORMANCE MEASURES USED: Number of flights and number of flight hours required to complete training in aircraft.

KEY FINDINGS: The fly-only group required more flight time to complete aircraft training than the other groups: the group trained with both visual and motion required less flying time.

REFERENCE: Browning, Ryan, and Scott (1978)

PRIMARY STUDY OBJECTIVE: Assess training effectiveness of Device 2F87F (Operational Flight Trainer for P-3 aircraft).

SUBJECT POPULATION: Aviators who had completed undergraduate multi-engine training in the S-2 aircraft, and who had standard instrument ratings. Twenty-seven aviators received training in Device 2F87F and in the aircraft; 68 received training only in the aircraft.

TRANSFER AIRCRAFT

- Type: Navy Fixed Wing
- Designation: P-3 Orion

FLIGHT SIMULATOR CHARACTERISTICS

- Name: 2F87F Operational Flight Trainer
- Motion System: 6 df Motion Platform
- Visual System: Camera-Modelboard, FOV is 38-degrees vertical and 50-degrees horizontal

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Landings, basic airwork, procedures for takeoff, instrument flight, and emergencies.

PERFORMANCE MEASURES USED: Instructor grades and proficiency ratings.

KEY FINDINGS: Simulator trained aviators reached proficiency in the aircraft in fewer hours than the aviators who received only aircraft training (8.6 vs 15.1 hours). Also, simulator trained aviators reached proficiency on landing tasks in fewer practice repetitions than aviators who received only aircraft training (17 vs 50 practice iterations).

REFERENCE: Byrum (1978)

PRIMARY STUDY OBJECTIVE: Evaluate the training effectiveness of a computer generated night visual system added to a conventional UH-1 Flight Simulator (instrument flight training simulator).

SUBJECT POPULATION: Trainees in the Army's Initial Entry Rotary-Wing Course who had no previous training in night flying.

TRANSFER AIRCRAFT

- Type: Army Rotary Wing
- Designation: UH-1 (Huey)

FLIGHT SIMULATOR CHARACTERISTICS

- Name: UH-1 Instrument Flight Simulator (equipped with a prototype computer generated night visual system develop by Singer-Link).
- Motion System: 5 df motion platform
- Visual System: Computer generated night scenes; "narrow" FOV (values not given).

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Night takeoff and climb, night cruise, night approach, night autorotation, and instrument approach and breakout to landing.

PERFORMANCE MEASURES USED: Instructor ratings and trials to criterion.

KEY FINDINGS: No measurable training transfer from simulator to aircraft.

REFERENCE: Edwards, Weyer, and Smith (1978)

PRIMARY STUDY OBJECTIVE: Validate a visual discrimination pretraining program designed to facilitate learning of the final turn to a landing approach.

SUBJECT POPULATION: Thirty-eight flight students who had previous experience in the T-41 aircraft (average of 19 hours), the T-4 simulator (average of 5 hours), and the T-37 aircraft (average of 7 rides).

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: N/A
- Motion System: N/A
- Visual System: A series of photographs showing extra-cockpit display scenes for normal and three magnitudes of errors in altitude and flight path.

INDEPENDENT VARIABLES INVESTIGATED (if any): type of pretraining: none, cognitive pretraining on procedures and parameters only, visual discrimination pretraining, training in the Advanced Simulator For Pilot Training (ASPT), and both visual discrimination pretraining, and training in the ASPT.

TASKS TRAINED IN FLIGHT SIMULATOR: Final turn to landing approach.

PERFORMANCE MEASURES USED: Instructor ratings and recorded deviations for optimal flight path in ASPT.

KEY FINDINGS: No apparent transfer from visual discrimination pretraining to aircraft.

REFERENCE: Dohme and Millard (1985)

PRIMARY STUDY OBJECTIVE: Compare the relative effectiveness of Primary training in the TH-55 helicopter with Primary training in the AH-1 Flight Weapons Simulator (FWS). (The purpose of Primary training is to prepare Initial Entry Rotary Wing (IERW) students for UH-1 Transition training.)

SUBJECT POPULATION: Twenty Army IERW students with no prior flying experience; 10 received training in the AH FWS and 10 received conventional training in the TH-55 aircraft.

TRANSFER AIRCRAFT

- Type: Army Rotary Wing
- Designation: UH-1 (Huey)

FLIGHT SIMULATOR CHARACTERISTICS

- Name: AH-1 Flight Weapons Simulator (FWS) (programmed to fly as much as possible like the UH-1 aircraft).
- Motion System: 6 df motion platform
- Visual System: Camera-Modelboard; FOV is 36-degree vertical and 101-degree horizontal (two windows).

INDEPENDENT VARIABLES INVESTIGATED (if any): Motion vs no-motion.

TASKS TRAINED IN FLIGHT SIMULATOR: The simulator trained students followed essentially the same training syllabus as the TH-55 trained students. The tasks include: before-takeoff and after-landing checks, hovering tasks, takeoffs and landings, traffic patterns, emergency procedures, and other air work.

PERFORMANCE MEASURES USED: Training grades during Primary and throughout UH-1 Transition (daily grades and end-of-phase grades), number of setbacks, and maneuver scores on out-of-curriculum UH-1 checkride.

KEY FINDINGS: Training in the AH1CWS transferred to the UH-1 aircraft as well as training in the TH-55 aircraft. An even greater degree of transfer would be expected from a flight simulator specifically designed for the UH-1 aircraft. No evidence suggested that platform motion benefitted training.

REFERENCE: Flexman, Roscoe, Williams, and Williges (1972)

PRIMARY STUDY OBJECTIVE: Assess the benefits of low-fidelity visual feedback on acquisition of traffic pattern flight.

SUBJECT POPULATION: Students with no prior flight experience.

TRANSFER AIRCRAFT

- Type: Navy Fixed Wing
- Designation: SNJ-4

FLIGHT SIMULATOR CHARACTERISTICS

- Name: 1-CA-2 SNJ Link Trainer
- Motion System: 2 df Motion Platform
- Visual System: Stationary picture of ground and horizon line; instructor traced approximate flight path on chalk board that was visible to students.

INDEPENDENT VARIABLES INVESTIGATED (if any): None

TASKS TRAINED IN FLIGHT SIMULATOR: Traffic pattern flight.

PERFORMANCE MEASURES USED: Trials and time to reach criterion in aircraft; errors.

KEY FINDINGS: Positive transfer of training demonstrated.

REFERENCE: Gray and Fuller (1977)

PRIMARY STUDY OBJECTIVE: Assess the training transfer from T-37 configured research simulator to the F-5B aircraft.

SUBJECT POPULATION: Graduates from the Air Force Undergraduate Pilot Training program; 250 to 275 flight hours.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: F-5B

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Advance Simulator for Pilot Training (ASPT) configured for T-37 aircraft.
- Motion System: 6 df Motion Platform
- Visual System: Computer generated; FOV is 150-degrees vertical and 300-degrees horizontal

INDEPENDENT VARIABLES INVESTIGATED (if any): Student aptitude; presence vs. absence of motion during simulator training.

TASKS TRAINED IN FLIGHT SIMULATOR: Bombing runs.

PERFORMANCE MEASURES USED: Circular error for 10-degree, 15-degree, and 30-degree dives; number of qualifying bombs; and instructor ratings of performance.

KEY FINDINGS: Positive training transfer demonstrated. Amount of transfer independent of student aptitude and presence/absence of motion.

REFERENCE: Hagin (1976)

PRIMARY STUDY OBJECTIVE: The assessment of effect on training transfer of motion/no motion, g-seat/no g-seat, and narrow vs wide FOV (three separate studies).

SUBJECT POPULATION: Eight student pilots with no previous T-37 flight experience (Study 1); three experienced T-37 instructor pilots (Study 2); eight student pilots with no previous T-37 experience (Study 3)

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Advanced Simulator for Pilot Training (ASPT)
- Motion System: 6 df Motion Platform
- Visual System: Computer generated; FOV is 150-degrees vertical and 300-degrees horizontal

INDEPENDENT VARIABLES INVESTIGATED (if any): Motion vs. no motion; g-seat vs. no g-seat; and narrow vs. full FOV.

TASKS TRAINED IN FLIGHT SIMULATOR: Basic airwork, takeoffs, and landings (Study 1); takeoffs, landings, aileron rolls, and slow flight (Study 2); all tasks in the USAF undergraduate syllabus (Study 3).

PERFORMANCE MEASURES USED: Instructor ratings, task iterations to proficiency, mean of automated performance measures, and hours to proficiency in aircraft.

KEY FINDINGS: Simulator performance (experienced instructor pilots) was superior with wide FOV. No evidence that training transfer was increased by presence of platform motion or g-seat motion during simulator training. Transfer of training was demonstrated.

REFERENCE: Holman (1979) STUDY 1: Aircraft Qualification Training

PRIMARY STUDY OBJECTIVE: Evaluate effectiveness of CH-47 flight simulator for Aircraft Qualification Training (AQC).

SUBJECT POPULATION: Twenty-four AQC students in experimental group (simulator/aircraft trained); 35 AQC students in control group (aircraft trained).

TRANSFER AIRCRAFT

- Type: Army Rotary Wing
- Designation: CH-47 (Chinook)

FLIGHT SIMULATOR CHARACTERISTICS

- Name: CH-47 Flight Simulator
- Motion System: 6 df Motion Platform
- Visual System: Camera-modelboard for forward window and computer generated for chin window; FOV of forward window is 36-degree vertical and 48-degree horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: 32 tasks; takeoffs and landings, airwork sling load operations, emergency procedures, etc.

PERFORMANCE MEASURES USED: Instructor pilot ratings (12-point), trials to criterion, time to criterion, cumulative transfer effectiveness ratios (trials, time and combined).

KEY FINDINGS: Positive transfer demonstrated to most but not all tasks investigated. Low transfer typically found for tasks performed close to the ground at slow speed (e.g., hovering maneuvers, shallow approaches, confined area operations, and external load operations). Low transfer may be due to limited FOV, low fidelity handling qualities at slow speeds, low fidelity motion cueing, or a combination of the three factors.

REFERENCE: Holman (1979) STUDY 2; Continuation Training

PRIMARY STUDY OBJECTIVE: Evaluate effectiveness of CH-47 flight simulator for training FORSCOM aviators already qualified in the CH-47 aircraft.

SUBJECT POPULATION: Twenty-eight FORSCOM aviators qualified and current in the CH-47 aircraft (15 experimental and 13 control group).

NOTES ON PROCEDURES: During a six-month study period, both the experimental- and control -group aviators flew mission-support missions in the aircraft, but were not permitted to spend any time flying the aircraft for the sole purpose of individual or crew training. The experimental- and control-group aviators accumulated an average of 45.2 and 58.0 aircraft hours, respectively. The experimental-group aviators accumulated a total of 29.7 hours in the CH-47FS during the test period.

TRANSFER AIRCRAFT

- Type: Army Rotary Wing
- Designation: CH-47 (Chinook)

FLIGHT SIMULATOR CHARACTERISTICS

- Name: CH-47 Flight Simulator
- Motion System: 6 df Motion Platform
- Visual System: Camera-modelboard for forward window and computer generated for chin window; FOV is 36-degree vertical and 48-degree horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Twenty-four tasks; takeoffs and landings, airwork, sling load operations, emergency procedures, etc.

PERFORMANCE MEASURES USED: Instructor pilot ratings (12-point) of individual maneuvers on pretest and posttest checkrides; overall test scores (sum of maneuver scores X maneuver-difficulty rating).

KEY FINDINGS: The pretest and posttest mean scores of the control group did not differ significantly, indicating that the mission-support flying was sufficient to maintain skills. The mean posttest scores were higher than the mean pretest scores for the experimental group, indicating that the use of the simulator to augment aircraft flying significantly improved aviator's performance.

REFERENCE: Jacobs and Roscoe (1975)

PRIMARY STUDY OBJECTIVE: Assess effectiveness of GAT-2 Simulator training for training novice students to fly Piper Cherokee aircraft.

SUBJECT POPULATION: Students with no prior flight experience.

TRANSFER AIRCRAFT

- Type: Civil Fixed Wing
- Designation: Piper Cherokee

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Singer-Link General Aviation Training (GAT-2)
- Motion System: 2 df Motion Platform
- Visual System: none

INDEPENDENT VARIABLES INVESTIGATED (if any): Presence vs. absence of motion; normal vs. random direction banking motion.

TASKS TRAINED IN FLIGHT SIMULATOR: Eleven basic flight tasks (tasks not specified).

PERFORMANCE MEASURES USED: Trials to criterion.

KEY FINDINGS: Positive transfer demonstrated. Amount of transfer greater for normal than for random direction of banking motion; normal-motion and no-motion groups did not differ in amount of training transfer.

REFERENCE: Lintern and Roscoe (1978)

PRIMARY STUDY OBJECTIVE: Assess effect of supplementary visual cues on training transfer.

SUBJECT POPULATION: Novice aviators.

TRANSFER AIRCRAFT

- Type: Civil Fixed Wing
- Designation: Piper Cherokee

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Singer-Link General Aviation Trainer (GAT-2)
- Motion System: 2 df Motion Platform
- Visual System: Computer generated display with limited detail in scene.

INDEPENDENT VARIABLES INVESTIGATED (if any): Presence/absence of supplementary visual cues.

TASKS TRAINED IN FLIGHT SIMULATOR: Approaches and landings.

PERFORMANCE MEASURES USED: Instructor ratings; number of unassisted landings.

KEY FINDINGS: Positive transfer demonstrated.

REFERENCE: Martin and Waag (1977a)

PRIMARY STUDY OBJECTIVE: Assess the effect of platform motion on the transfer of basic contact flight skills trained in simulator.

SUBJECT POPULATION: Undergraduate student aviators with between 13 and 80 aircraft hours.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Advanced Simulator for Pilot Training (ASPT)
- Motion System: 6 df Motion Platform
- Visual System: Computer Generated; FOV is 150-degrees vertical and 300-degrees horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): Presence/absence of motion.

TASKS TRAINED IN FLIGHT SIMULATOR: Takeoff, overhead patterns, approach and landing and slow flight.

PERFORMANCE MEASURES USED: Instructor ratings.

KEY FINDINGS: Positive transfer demonstrated for all maneuvers. No evidence that motion influenced training transfer.

REFERENCE: Martin and Waag (1977b)

PRIMARY STUDY OBJECTIVE: Assess the effect of platform motion on the training transfer of aerobatics training in the simulator.

SUBJECT POPULATION: No prior flight experience in T-37 aircraft.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Advanced Simulator for Pilot Training (ASPT)
- Motion System: 6 df Motion Platform
- Visual System: Computer Generated; FOV is a 150-degree vertical and 300-degree horizontal

INDEPENDENT VARIABLES INVESTIGATED (if any): Presence/absence of motion.

TASKS TRAINED IN FLIGHT SIMULATOR: Aileron roll, split S, loop, lazy 8, Immelman, bank and roll, Cuban 8, and clover leaf.

PERFORMANCE MEASURES USED: Instructor ratings.

KEY FINDINGS: Statistically significant transfer found for only one maneuver: bank and roll. No evidence that motion influenced transfer.

REFERENCE: Payne, Hirsh, Semple, Farmer, Spring, Sanders, Wimer, Carter, and Hu (1976)

PRIMARY STUDY OBJECTIVE: Assess transfer of simulator training on air-to-air combat.

SUBJECT POPULATION: Graduates from Undergraduate Pilot Training with about 350 aircraft hours; experienced aviators with more than 1220 hours of aircraft hours.

TRANSFER AIRCRAFT

- Type: Navy Fixed Wing
- Designation: F-4J

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Northrop Air-to-Air Combat Simulator
- Motion System: Yes
- Visual System: Visual projection of earth, sky and adversary aircraft; FOV is 210-degrees.

INDEPENDENT VARIABLES INVESTIGATED (if any): Aviator experience level.

TASKS TRAINED IN FLIGHT SIMULATOR: Lag pursuit, lag roll, high yo yo, low yo yo, barrel roll attack, rolling scissors, head-on maneuver and guns defense.

PERFORMANCE MEASURES USED: Instructor ratings; measure of final position after engagement.

KEY FINDINGS: Positive transfer for all tasks was indicated by the metric "final position after engagement"; instructor ratings showed positive transfer only for rolling scissors maneuver.

REFERENCE: Pohlman and Reed (1978)

PRIMARY STUDY OBJECTIVE: Assess the effect of platform motion on the transfer of air-to-air combat skills trained in simulator.

SUBJECT POPULATION: Aviators undergoing F-4 transition training.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: F-4

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Simulator for Air-to-Air Combat
- Motion System: 6 df Motion Platform
- Visual System: Computer generated terrain image; camera model adversary aircraft image; FOV is 150-degrees vertical and 296-degrees horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): Presence/absence of motion.

TASKS TRAINED IN FLIGHT SIMULATOR: Acceleration maneuver, high yo yo, quarter plane, barrel roll attack, Immelman attack, log roll, separation, tactical formation, step up on perch, and defense maneuvers.

PERFORMANCE MEASURES USED: Instructor Ratings.

KEY FINDINGS: No positive transfer demonstrated, possibly because the students were not given instruction during simulator training. No evidence that motion influenced transfer in any way.

REFERENCE: Reed and Reed (1978)

PRIMARY STUDY OBJECTIVE: Assess transfer from training in an air refueling trainer.

SUBJECT POPULATION: Aviators undergoing F-4C qualification training.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: F-4C

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Air Refueling Director Lights Trainer
- Motion System: none
- Visual System: Dynamic presentation of receiver director lights on the underside of an air refueling tanker.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Air refueling

PERFORMANCE MEASURES USED: Instructor ratings

KEY FINDINGS: Positive transfer demonstrated, but only on the first training mission in the aircraft.

REFERENCE: Reid and Cyrus (1974)

PRIMARY STUDY OBJECTIVE: Assess transfer of simulator training on formation flying.

SUBJECT POPULATION: Aviators with an average of 82.5 hours in the T-37 aircraft and 30 hours in the T-38 aircraft. Seventy-two aviators served as subjects in Study 1; 48 aviators served as subjects in Study 2.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-38

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Formation Flight Trainer
- Motion System: none
- Visual System: Wide angle projected TV picture of lead aircraft.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Formation flight

PERFORMANCE MEASURES USED: Check Section aviator ratings (Study 1) and ratings by specially trained Instructor Pilots (Study 2).

KEY FINDINGS: Positive transfer demonstrated.

REFERENCE: Ryan, Scott, and Browning (1978)

PRIMARY STUDY OBJECTIVE: Assess the transfer of simulator training on landings (Study 1) and assess the effect of motion on training transfer (Study 2).

SUBJECT POPULATION: Ninety-five first tour Naval aviators (Study 1); 50 first tour Naval aviators, 39 motion and 11 no-motion (Study 2).

TRANSFER AIRCRAFT

- Type: Navy Fixed Wing
- Designation: P-3 Orion

FLIGHT SIMULATOR CHARACTERISTICS

- Name: 2F87F
- Motion System: 6 df Motion Platform
- Visual System: Camera-Modelboard; FOV is 38-degrees vertical and 50-degrees horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): Blocked simulator trials vs interspersed simulator and aircraft trails; motion vs no-motion.

TASKS TRAINED IN FLIGHT SIMULATOR: Final approach and landing.

PERFORMANCE MEASURES USED: Instructor ratings, flight hours to criterion, and landings to proficiency.

KEY FINDINGS: Positive transfer demonstrated for all simulator-trained groups. blocked simulator trials resulted in greater transfer than interspersed trials by authors attributed differences to methodological factors. Transfer was not influenced by presence/absence of motion, but every subject preferred the motion to the no-motion condition.

REFERENCE: Smith, Waters, and Edwards (1975)

PRIMARY STUDY OBJECTIVE: Assess the extent to which training on a multi-media cognitive pretraining package transfers to the aircraft.

SUBJECT POPULATION: Thirty undergraduate pilot students who had previous flight time in the T-41 aircraft.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Multi-media cognitive pretraining package
- Motion System: N/A
- Visual System: Films (8mm) and slides (35mm) taken from T-37 cockpit during approach and landing.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Overhead patterns

PERFORMANCE MEASURES USED: Number of segments and landmarks recognized; trials-to-criterion on the aircraft.

KEY FINDINGS: The cognitive pretraining instructional material produced consistently superior student pilot performance on both written tests and on inflight transfer of training evaluations.

REFERENCE: Thorpe, Varney, McFadden, LeMaster, and Short (1978)

PRIMARY STUDY OBJECTIVE: Determine the relative effectiveness of three types of simulator visual systems in KC-135 combat crew training.

SUBJECT POPULATION: Thirty recent graduates of undergraduate pilot training who were transitioning into the KC-135 copilot position.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: KC-135

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Boeing 707 Flight Simulator
- Motion System: 3 df Motion Platform
- Visual System: Day/night color computer image generation system, night only point light source computer image generation system, and camera-modelboard system.

INDEPENDENT VARIABLES INVESTIGATED (if any): Type of extra-cockpit display.

TASKS TRAINED IN FLIGHT SIMULATOR: Approach and landing.

PERFORMANCE MEASURES USED: Instructor ratings, number of aviators reaching proficiency in the simulator, total simulator time, and total number of successful landings.

KEY FINDINGS: The two computer generated displays were superior to the camera-modelboard system. However, significant transfer was demonstrated for all three visual systems. The amount of transfer cannot be assessed accurately because no aircraft-only control group was used.

REFERENCE: Woodruff and Smith (1974)

PRIMARY STUDY OBJECTIVE: Investigate the utility of an AF 37A-T4G simulator in Air Force undergraduate pilot training.

SUBJECT POPULATION: Air Force undergraduate pilot students with little or no flying experience.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: T-4G
- Motion System: 2 df Motion Platform
- Visual System: Film base visual system; FOV is 95-degrees vertical and 300-degrees horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Basic contact flight tasks and basic instrument flight; (Training sequence: simulator contact, aircraft contact, simulator instruments, and aircraft instruments.)

PERFORMANCE MEASURES USED: Aircraft hours required to reach criterion performance.

KEY FINDINGS: Significant transfer of training demonstrated. Contact training hours in the aircraft were reduced by about 20%; instrument hours in the aircraft were reduced by about 45%.

REFERENCE: Woodruff, Smith, Fuller, and Weyer (1976)

PRIMARY STUDY OBJECTIVE: Assess the transfer of training in the Advanced Simulator for Pilot Training on basic and advanced contact flight tasks.

SUBJECT POPULATION: Students in Air Force undergraduate pilot training program who had less than 50 hours of flight experience.

TRANSFER AIRCRAFT

- Type: Air Force Fixed Wing
- Designation: T-37

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Advanced Simulator for Pilot Training
- Motion System: 6 df Motion Platform
- Visual System: Computer generated display; FOV is 95-degrees vertical and 300-degrees horizontal.

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Basic and advanced contact flight tasks.

PERFORMANCE MEASURES USED: Time to reach criterion in the aircraft.

KEY FINDINGS: Substantial positive transfer demonstrated for basic contact flight, small amount of positive transfer demonstrated for advanced contact flight. However, due to scheduling difficulties, simulator training time was devoted to advanced contact flight.

REFERENCE: Young, Jensen, and Treschel (1973)

PRIMARY STUDY OBJECTIVE: Assess the effectiveness of simulator training using a very low fidelity visual system.

SUBJECT POPULATION: Students with little or no flight experience.

TRANSFER AIRCRAFT

- Type: Unknown
- Designation: Unknown

FLIGHT SIMULATOR CHARACTERISTICS

- Name: Unknown
- Motion System: Unknown
- Visual System: Runway and colored horizon

INDEPENDENT VARIABLES INVESTIGATED (if any): none

TASKS TRAINED IN FLIGHT SIMULATOR: Approaches and landings

PERFORMANCE MEASURES USED: Instructor ratings

KEY FINDINGS: Poor instruction precludes meaningful conclusions. Too little time devoted to simulator training. In addition, the visual system disappeared at flare point.